International Journal of Forecasting 33 (2017) 416–432

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Density forecast evaluation in unstable environments

Gloria González-Rivera[a,*], Yingying Sun[b]

[a] Department of Economics, University of California, Riverside, CA, USA

[b] Bank of the West, San Francisco, CA, USA

ARTICLE INFO

Keywords:
Generalized autocontour-based testing
Structural breaks
Phillips curve

ABSTRACT

We propose a density forecast evaluation method in the presence of instabilities, which are defined as breaks in any conditional moment of interest and/or in the functional form of the conditional density of the process. Within the framework of the autocontour-based tests proposed by González-Rivera et al. (2011) and González-Rivera and Sun (2015), we construct Sup- and Ave-type tests, calculated over a collection of subsamples in the evaluation period. These tests have asymptotic distributions that are nuisance-parameter free and they are correctly sized and very powerful for detecting breaks in the parameters of the conditional mean and conditional variance. A power comparison with the tests of Rossi and Sekhposyan (2013) shows that our tests are more powerful across the models considered in their work. We analyze the stability of a dynamic Phillips curve and find that the best one-step-ahead density forecast of changes in inflation is generated by a Markov switching model that allows state shifts in the mean and variance of inflation changes as well as in the coefficient that links inflation and unemployment.

1. Introduction

In general, instability refers to changes in the parameters of a proposed forecasting model over the forecasting horizon. For clarification purposes, consider a simple model $y_{t+1} = \beta' x_t + \sigma \varepsilon_{t+1}$, with $\varepsilon_t \sim i.i.d. N(0, 1)$. The model is unstable over time if the slope coefficients β can change over the forecasting sample, either smoothly or abruptly, to contain one or more breaks. We may also entertain a time-varying variance such that σ is also subject to breaks, and we may have different conditional probability density functions, e.g., more or less thick tails, in different periods. This definition is general enough to account for most of the types of instability that are discussed in the current applied econometric literature. To date, the most comprehensive survey of the subject is that provided by Rossi (2014) in the *Handbook of Economic Forecasting*, which reports extensive empirical evidence of instabilities in macroeconomic and financial data. Some examples follow.

The instability of predictive regressions, in which the significance of predictive regressors varies over different subsamples, has been documented in studies of the predictability of stock returns (see Goyal & Welch, 2003; Paye & Timmermann, 2006; Rapach & Zhou, 2014), in exchange rate predictions (see Rogoff & Stavrakeva, 2008; Rossi, 2006) and in output growth and inflation forecasts (see Rossi & Sekhposyan, 2010; Stock & Watson, 2003). Naturally, linked to this evidence is the econometric issue of testing for parameter stability and structural breaks in the data, which has an illustrious history. From the Chow (1960) test to more recent works such as those of Andrews (1993), Andrews and Ploberger (1994) and Pesaran and Timmermann (2002), among others, testing for breaks has focused mainly on the behavior of the conditional mean. This paper aims to extend the testing for instabilities to

* Corresponding author.
E-mail address: gloria.gonzalez@ucr.edu (G. González-Rivera).

http://dx.doi.org/10.1016/j.ijforecast.2016.10.003
0169-2070/

the *full* conditional density forecast that includes not only any conditional measure of interest (e.g., mean, variance, duration, etc.), but also the functional form of the assumed conditional density function. To the best of our knowledge, the literature on this question is very thin. We know only of the work by Rossi and Sekhposyan (R&S) (2013), who also tested for the instability of the density model, but using a different methodology. A comparison of the two approaches will be provided later.

The testing methodology that we propose is based on the AutoContouR (ACR) device introduced by González-Rivera, Senyuz, and Yoldas (2011) and González-Rivera and Yoldas (2012), and later generalized by González-Rivera and Sun (2015). A brief summary of this latter work follows. The null hypothesis is a correctly specified density forecast (with the joint hypothesis of correct dynamics in the moments of interest and a correct functional form of the density). We calculate the Rosenblatt (1952) probability integral transforms (PIT) that are associated with the point forecasts. Under the null, the PITs must be i.i.d uniformly distributed U[0,1]. The generalized autocontour (G-ACR) is a device (set of points) that is very sensitive to departures from the null in either direction; consequently, it provides the basis for very powerful tests. More specifically, for a time series of PITs, we construct the G-ACRs as squares (in the univariate case) of different probability areas within the maximum square (area of 1), or as hyper-cubes (in the multivariate case) of different probability volumes within the maximum hyper-cube formed by a multidimensional uniform density $[0, 1]^n$ for $n \geq 2$. By statistical comparisons of the location of the empirical PITs and the volume of the empirical G-ACRs with the location and volume of the population G-ACRs, we are able to construct a variety of tests for correct density forecasts. Since the shapes of the G-ACRs can be visualized, we can extract information about where the rejection of the null hypothesis comes from, and how. This testing framework is the foundation of the new stability tests, the Sup- and Ave-type statistics, proposed in this paper. In a potentially unstable data environment, we form rolling subsamples within the forecasting sample. For every subsample, we apply a battery of G-ACR tests and construct Sup- and Ave-type statistics for detecting instabilities. Although the limiting distribution of these tests is a function of Brownian motions, the tests are nuisance-parameter free and their distributions can be tabulated.[1]

The R&S (2013) tests and our tests have similar null hypotheses, i.e., the constancy of the density model over the prediction sample, although their statistics allow for dynamic misspecification. The R&S tests follow the setup of Corradi and Swanson (2006), so that their tests, which are also based on the PITs $\{u_t\}$ of the proposed model, are a function of the distance between the empirical cumulative distribution function and that of the uniform distribution, which is a 45 degree line. Under dynamic misspecification, the PITs are still distributed uniformly in [0,1], but are no longer independent. Thus, the R&S tests borrow from Corradi and Swanson, in that the limiting variance of the statistics has to take into account the potential lack of independence. Their κ_P test is a Kolmogorov–Smirnov-type statistic, and their C_P is a Cramer-von Mises-type statistic, which mainly exploits the "identical distribution" property of the PITs under the null. Our G-ACR tests are based on the object "autocontour", and measure the independence, denseness and uniform distribution of the PITs over a collection of squares in a two-dimensional space (u_{t-k}, u_t). By construction, our tests exploit the independence properties of the PITs directly. The asymptotic distribution of the R&S tests is based on the statistical properties of an empirical process. Our tests are simpler, in that they rely on the properties of a binary indicator with well-defined moments. When the parameter uncertainty is non-negligible, the critical values of the R&S tests and the G-ACR tests are obtained by simulation.

The paper is organized as follows. Section 2 reviews the G-ACR approach in order to make the exposition self-contained, and introduces the new statistics with their asymptotic distributions. Section 3 assesses the finite sample properties (size and power) of the tests. We offer an extensive assessment by considering (i) fixed, rolling, and recursive estimation schemes; (ii) different ratios of prediction to estimation sample sizes; and (iii) break points that occur in different periods of the prediction sample. We also present a comparison of the power of our tests with those of Rossi and Sekhposyan (2013). Section 4 uses the tests to assess the stability of the Phillips curve from 1958 onwards by evaluating the models proposed by Amisano and Giacomini (2007). Section 5 concludes. Appendix A contains mathematical proofs and Appendix B describes the parametric bootstrap for correcting the size of the tests. We also provide a supplementary file with additional simulation materials (see Appendix C).

2. Statistics and asymptotic distributions

2.1. Construction of the statistics

The test statistics are based on the autocontour (ACR) and generalized autocontour (G-ACR) methodologies proposed by González-Rivera et al. (2011), González-Rivera and Yoldas (2012), and González-Rivera and Sun (2015), which provide powerful tests for the dynamic specification of the conditional density model in either in-sample or out-of-sample environments. In the present context, we adapt these tests to instances where instabilities may be present in the data, so that we will also be able to detect unstable periods beyond the evaluation of the density model.

Let Y_t denote the random process of interest with a conditional density function $f(y_t|\Omega_{t-1})$, where Ω_{t-1} is the information set available up to time $t-1$. Observe that the random process Y_t could enjoy very general statistical properties, e.g., heterogeneity, dependence, etc. The researcher will construct the conditional model by specifying a conditional mean, conditional variance or other conditional moments of interest, and making distributional assumptions as to the functional form of $f(\cdot)$. Based on the conditional model, she will then proceed

[1] Although we focus on out-of-sample density forecasts, the methodology proposed in this paper can also be applied to in-sample specification testing.

to construct a density forecast. If the proposed predictive density model for Y_t, i.e., $\{f_t^*(y_t|\Omega_{t-1})\}_{t=1}^T$, coincides with the true conditional density $\{f_t(y_t|\Omega_{t-1})\}_{t=1}^T$, then the sequence of probability integral transforms (PITs) of $\{Y_t\}_{t=1}^T$ w.r.t $\{f_t^*(y_t|\Omega_{t-1})\}_{t=1}^T$ (i.e., $\{u_t\}_{t=1}^T$) must be *i.i.d.* $U(0,1)$, where $u_t = \int_{-\infty}^{y_t} f_t^*(v_t|\Omega_{t-1})dv_t$. Thus, the null hypothesis $H_0 : f_t^*(y_t|\Omega_{t-1}) = f_t(y_t|\Omega_{t-1})$ is equivalent to the null hypothesis $H_0' : \{u_t\}_{t=1}^T$ is *i.i.d.* $U(0,1)$ (see Diebold, Gunther, & Tay, 1998; Rosenblatt, 1952).

In order to provide a self-contained exposition, we briefly review the approach of González-Rivera and Sun (2015). We start by constructing the generalized autocontours (G-ACR) under i.i.d. uniform PITs. Under H_0' : $\{u_t\}_{t=1}^T i.i.d.\ U(0,1)$, the G-ACR$_{\alpha_i,k}$ is defined as the set $B(\cdot)$ of points in the plane (u_t, u_{t-k}), such that the square with $\sqrt{\alpha_i}$ sides contains at most α_i% of observations, i.e.,

$$\text{G-ACR}_{\alpha_i,k} = \{B(u_t,u_{t-k}) \subset \Re^2 \parallel 0 \le u_t \le \sqrt{\alpha_i} \text{ and } 0 \le u_{t-k} \le \sqrt{\alpha_i},\ s.t.: u_t \times u_{t-k} \le \alpha_i\}.$$

We construct an indicator series I_t^{k,α_i} as follows:

$$I_t^{k,\alpha_i} = \mathbf{1}((u_t,u_{t-k}) \subset \text{G-ACR}_{\alpha_i,k}) = \mathbf{1}(0 \le u_t \le \sqrt{\alpha_i}, 0 \le u_{t-k} \le \sqrt{\alpha_i}).$$

Based on this indicator, González-Rivera and Sun (2015) proposed the following t-tests and chi-square statistics for testing the null hypothesis H_0' : $\{u_t\}_{t=1}^T i.i.d.\ U(0,1)$, which is equivalent to testing for the correct specification of the full conditional density model.

(1) t-ratio testing: for a fixed autocontour α_i and fixed lag k,[2]

$$z \equiv \frac{\sqrt{T-k}(\widehat{\alpha}_i - \alpha_i)}{\sigma_{k,i}} \xrightarrow{d} N(0,1),$$

where $\widehat{\alpha}_i = \frac{\sum_{t=k+1}^T I_t^{k,\alpha_i}}{T-k}$, and $\sigma_{k,i}^2$ is the asymptotic variance of $\widehat{\alpha}_i$.

(2) Chi-squared testing:

(2.1) For a fixed lag k, $\mathbf{C}_k'\Omega_k^{-1}\mathbf{C}_k \xrightarrow{d} \chi_C^2$ where $\mathbf{C}_k = (c_{k,1},\ldots,c_{k,C})'$ is a $C \times 1$ stacked vector with elemen t $c_{k,i} = \sqrt{T-k}(\widehat{\alpha}_i - \alpha_i)$, and Ω_k the asymptotic variance–covariance matrix of the vector $\mathbf{C}_k$.

(2.2) For a fixed autocontour α_i, $\mathbf{L}_{\alpha_i}'\Lambda_{\alpha_i}^{-1}\mathbf{L}_{\alpha_i} \xrightarrow{d} \chi_K^2$, where $\mathbf{L}_{\alpha_i} = (\ell_{1,\alpha_i},\ldots\ell_{K,\alpha_i})'$ is a $K \times 1$ stacked vector with element $\ell_{k,\alpha_i} = \sqrt{T-k}(\widehat{\alpha}_i - \alpha_i)$, and Λ_{α_i} is the asymptotic variance–covariance matrix of the vector $\mathbf{L}_{\alpha_i}$.

The contribution of this work is its consideration of potentially unstable environments. We propose a new battery of tests that will be calculated within the following rolling subsample scheme. The total sample size T is divided into two parts: in-sample observations (R) and out-of-sample observations (P). We form subsamples of size r from $t-r+1$ up to t, where $t = R+r,\ldots,T$, then evaluate the proposed predictive density within each subsample by calculating the three statistics reviewed in tests (1), (2.1), and (2.2), which we call z, C and L. As a result, we obtain three sets of $n \equiv T-r-R+1$ tests each, i.e., $\{z_j\}_{j=1}^n$, $\{C_j\}_{j=1}^n$ and $\{L_j\}_{j=1}^n$. Finally, we detect instabilities through the construction of Sup-type and Avg-type statistics, by taking the supremum (S) and the average (A) respectively over each set $\{|z_j|\}_{j=1}^n$, $\{C_j\}_{j=1}^n$ and $\{L_j\}_{j=1}^n$, to obtain the following six statistics: $S_{|z|}$, S_C, S_L and $A_{|z|}$, A_C, and A_L.

2.2. Asymptotic properties of the statistics

Under the following set of assumptions, we provide three propositions, for which proofs are found in Appendix A.

- A1: For $T \to \infty$, $R \to \infty$ and $P \to \infty$, $\lim_{T\to\infty} \frac{P}{R} = 0$; and for $r, P \to \infty$, $\lim_{T\to\infty} \frac{r-k}{P} = m$, where r is the size of the rolling subsample in the evaluation set, $m \in (0,1)$, and k is the lag in the indicator I_t^{k,α_i}.[3]
- A2: $E|I_t^{k,\alpha_i}|^q < \Delta < \infty$ for some $q \ge 2$. This assumption is trivial, since the second moments of the indicator are well defined, as we will see next.
- A3: The data $\{y_t\}$ come from a stationary and ergodic process.[4]

Proposition 1. *Let J be the index that identifies a particular subsample in the evaluation period, i.e., $J = [Ps]$, $s \in [m,1]$, $[Pm] = r-k$, and let $\widehat{\alpha}_i(J) = \frac{\sum_{t=R+1+J-r+k}^{R+J} I_t^{k,\alpha_i}}{r-k}$ be the corresponding subsample proportion based on the indicator. The Sup- and Avg-tests are*

$$S_{|z|} = \sup_J \left| \frac{\sqrt{r-k}(\widehat{\alpha}_i(J) - \alpha_i)}{\sigma_{k,\alpha_i}} \right|$$

$$A_{|z|} = \frac{1}{P-r+1} \sum_{\underline{J}}^{\bar{J}} \left| \frac{\sqrt{r-k}(\widehat{\alpha}_i(J) - \alpha_i)}{\sigma_{k,\alpha_i}} \right|,$$

where $\sigma_{k,\alpha_i}^2 = \alpha_i(1-\alpha_i) + 2\alpha_i^{3/2}(1-\alpha_i^{1/2})$.

[2] The lag k is chosen by the researcher. As k changes, we will have a sequence of t-ratios for $k = 1,2,\ldots,K$. We can plot these t-ratios as a function of k in order to obtain a graph that is similar in spirit to classical autocorrelograms, although it is actually a plot of t-ratios in our case. This graph is useful for the visual detection of dynamic misspecification. The portmanteau test L_{α_i} is a joint test that includes as many values of k as we wish to test. Including the sequence of lagged PITs in the construction of the tests is one important difference from other tests. This feature is one of the reasons why our tests are more powerful than most.

[3] According to several Monte Carlo studies (West, 2006, and the references therein), the ratio P/R is considered sufficiently small for the parameter uncertainty to be negligible when $P/R \le 10\%$. When $\lim_{T\to\infty} \frac{P}{R} \ne 0$, parameter uncertainty will affect the asymptotic variance σ_{k,α_i}^2 of the statistic. By taking a mean value expansion around the true parameter values θ_0 and applying Slutsky's theorem, we obtain

$$\sqrt{P}(\hat{\alpha}_i(\hat{\theta}) - \alpha_i) = \sqrt{P}(\hat{\alpha}_i(\theta_0) - \alpha_i) + \frac{\sqrt{P}}{\sqrt{R}}\sqrt{R}(\hat{\theta} - \theta_0)' \lim_{R\to\infty} E\left\{\frac{\partial \hat{\alpha}_i}{\partial \theta}|_{\theta=\theta_0}\right\} + o_p(1).$$

Both the variance of the second term of the expansion and the covariance between the first and second terms will contribute to the asymptotic variance of the statistics. For the calculation of these terms, see González-Rivera et al. (2011). In practice, these terms will be calculated using a fully parametric bootstrap procedure.

[4] This assumption can be relaxed to include more general mixing processes because the relevant conditions for invoking limiting theorems as the FCLT are those of the indicator process.

Table 1
Asymptotic distributions of the $S_{|z|}$ and $A_{|z|}$ statistics. The percentiles are obtained from 2000 replications, with a sample size of 20,000 observations in each replication.

	m								
Percentile	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Asymptotic distribution of $S_{\|z\|}$ statistics.									
99%	3.950	3.713	3.510	3.396	3.428	3.213	3.089	3.112	2.951
95%	3.502	3.267	3.066	2.926	2.866	2.679	2.537	2.439	2.406
90%	3.292	2.987	2.845	2.642	2.571	2.361	2.240	2.121	2.015
80%	3.020	2.684	2.522	2.372	2.217	2.053	1.935	1.849	1.655
70%	2.843	2.489	2.315	2.156	1.999	1.834	1.697	1.582	1.392
60%	2.702	2.343	2.145	1.983	1.819	1.653	1.529	1.383	1.197
50%	2.569	2.203	2.015	1.830	1.664	1.515	1.365	1.221	1.021
40%	2.457	2.090	1.884	1.684	1.515	1.373	1.212	1.066	0.867
30%	2.339	1.975	1.747	1.540	1.379	1.228	1.071	0.915	0.730
20%	2.201	1.822	1.586	1.384	1.226	1.074	0.923	0.791	0.606
10%	2.033	1.629	1.420	1.199	1.055	0.901	0.761	0.631	0.479
5%	1.922	1.492	1.290	1.065	0.943	0.813	0.663	0.535	0.394
1%	1.730	1.265	1.072	0.829	0.739	0.631	0.498	0.409	0.287
Asymptotic distribution of $A_{\|z\|}$ statistics.									
99%	1.241	1.435	1.656	1.824	2.204	2.181	2.360	2.542	2.583
95%	1.078	1.199	1.355	1.487	1.697	1.694	1.774	1.854	1.998
90%	1.004	1.088	1.206	1.300	1.418	1.406	1.490	1.586	1.666
80%	0.918	0.970	1.035	1.101	1.115	1.128	1.183	1.264	1.256
70%	0.870	0.891	0.922	0.948	0.939	0.928	0.966	1.017	1.019
60%	0.825	0.825	0.827	0.837	0.818	0.796	0.807	0.824	0.825
50%	0.789	0.760	0.760	0.741	0.710	0.685	0.682	0.675	0.661
40%	0.752	0.706	0.690	0.653	0.615	0.581	0.551	0.534	0.507
30%	0.711	0.649	0.617	0.568	0.524	0.487	0.453	0.417	0.378
20%	0.673	0.589	0.544	0.497	0.451	0.409	0.358	0.320	0.274
10%	0.608	0.517	0.465	0.409	0.359	0.319	0.269	0.234	0.186
5%	0.567	0.458	0.413	0.342	0.310	0.260	0.219	0.188	0.140
1%	0.478	0.381	0.325	0.251	0.235	0.200	0.152	0.135	0.096

Given assumptions A1–A3, and under the null hypothesis of *i.i.d.* $U(0,1)$ PITs, the asymptotic distribution of the tests, for $P \to \infty$, is as follows:

$$S_{|z|} \xrightarrow{d} \sup_{s\in[m,1]} \frac{|W(s)-W(s-m)|}{\sqrt{m}}$$

$$A_{|z|} \xrightarrow{d} \int_{\underline{s}}^{\bar{s}} \frac{|W(s)-W(s-m)|}{\sqrt{m}} ds,$$

where $W(\cdot)$ is a standard univariate Brownian motion.

Proposition 2. *For a fixed lag k, write* $c_{k,i}(J) = \sqrt{r-k}(\widehat{\alpha_i}(J)-\alpha_i)$ *and stack* $c_{k,i}(J)$ *for different autocontour levels* $i = 1, 2, \ldots, C$, *such that* $\mathbf{C}_k(J) = (c_{k,1}(J), \ldots, c_{k,C}(J))'$ *is the* $C \times 1$ *stacked vector. The Sup- and Avg-tests are*

$$S_C = \sup_J \mathbf{C}_k(J)'\Omega_k^{-1}\mathbf{C}_k(J)$$

$$A_C = \frac{1}{P-r+1}\sum_{\underline{J}}^{\bar{J}} \mathbf{C}_k(J)'\Omega_k^{-1}\mathbf{C}_k(J),$$

where Ω_k *is the asymptotic variance and covariance matrix of the random vector* $\mathbf{C}_k(J)$. *Given assumptions* A1–A3 *, and under the null hypothesis of i.i.d.* $U(0,1)$ *PITs, the asymptotic distribution of the tests, for* $P \to \infty$, *is*

$$S_C \xrightarrow{d} \sup_{s\in[m,1]} \frac{(\mathbf{W}(s)-\mathbf{W}(s-m))'(\mathbf{W}(s)-\mathbf{W}(s-m))}{m}$$

$$A_C \xrightarrow{d} \int_{\underline{s}}^{\bar{s}} \frac{(\mathbf{W}(s)-\mathbf{W}(s-m))'(\mathbf{W}(s)-\mathbf{W}(s-m))}{m} ds,$$

where $\mathbf{W}(\cdot)$ *is a standard C-variate Brownian motion.*

Proposition 3. *For a fixed autocontour* α_i, *write* $\ell_{k,\alpha_i}(J) = \sqrt{r-k}(\widehat{\alpha_i}(J)-\alpha_i)$ *and stack* ℓ_{k,α_i} *for* $k = 1, \ldots .K$. *Let* $\mathbf{L}_{\alpha_i}(J) = (\ell_{1,\alpha_i}(J), \ldots \ell_{K,\alpha_i}(J))'$ *be the* $K \times 1$ *stacked vector. The Sup- and Avg-tests are*

$$S_L = \sup_J \mathbf{L}_{\alpha_i}(J)'\Lambda_{\alpha_i}^{-1}\mathbf{L}_{\alpha_i}(J)$$

$$A_L = \frac{1}{P-r+1}\sum_{\underline{J}}^{\bar{J}} |\mathbf{L}_{\alpha_i}(J)'\Lambda_{\alpha_i}^{-1}\mathbf{L}_{\alpha_i}(J)|,$$

where Λ_{α_i} *is the asymptotic variance and covariance matrix for the random vector* $\mathbf{L}_{\alpha_i}$. *Given assumptions* A1–A3 *, and under the null hypothesis of i.i.d.* $U(0,1)$ *PITs, the asymptotic distribution of the tests, for* $P \to \infty$, *is*

$$S_L \xrightarrow{d} \sup_{s\in[m,1]} \frac{(\mathbf{W}(s)-\mathbf{W}(s-m))'(\mathbf{W}(s)-\mathbf{W}(s-m))}{m}$$

$$A_L \xrightarrow{d} \int_{\underline{s}}^{\bar{s}} \frac{(\mathbf{W}(s)-\mathbf{W}(s-m))'(\mathbf{W}(s)-\mathbf{W}(s-m))}{m} ds,$$

where $\mathbf{W}(\cdot)$ *is a standard K-variate Brownian motion.*

We tabulate the percentiles of the asymptotic distributions of the tests provided in Propositions 1–3. Since the distributions depend on m, which is the proportion of the rolling subsample to the total evaluation sample, we consider values of $m \in [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]$. Table 1 reports the percentiles of the distributions of

Table 2
Asymptotic distributions of the S_C and A_C statistics (13 autocontours).

	m								
Percentile	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Asymptotic distribution of S_C statistics.									
99%	42.488	39.592	38.534	37.694	37.361	35.925	36.051	34.591	31.804
95%	37.159	34.956	32.983	32.277	30.902	29.597	29.008	27.773	25.962
90%	35.155	32.562	30.722	29.378	28.069	27.176	26.181	24.685	23.356
80%	32.624	29.964	28.071	26.673	25.329	24.182	22.935	21.621	20.175
70%	30.932	28.330	26.359	24.906	23.421	22.132	20.932	19.423	18.125
60%	29.689	26.991	24.908	23.416	21.858	20.579	19.179	17.901	16.550
50%	28.540	25.798	23.648	22.111	20.626	19.136	17.708	16.501	15.241
40%	27.482	24.522	22.468	20.900	19.254	17.672	16.440	15.111	13.829
30%	26.508	23.385	21.197	19.510	17.873	16.413	15.058	13.685	12.477
20%	25.351	22.174	19.836	18.193	16.514	14.927	13.582	12.250	10.898
10%	23.731	20.400	18.067	16.333	14.568	13.061	11.741	10.480	8.987
5%	22.376	18.902	16.821	14.879	13.338	11.876	10.383	9.037	7.852
1%	20.394	16.833	14.737	12.807	11.208	9.884	8.275	6.929	5.905
Asymptotic distribution of A_C statistics.									
99%	16.342	18.610	20.265	22.301	24.658	26.534	27.422	28.202	27.987
95%	15.298	16.576	17.794	18.927	19.924	20.838	21.866	22.084	22.533
90%	14.762	15.734	16.456	17.287	18.022	18.629	18.996	19.612	19.965
80%	14.099	14.648	15.143	15.477	16.014	16.449	16.761	16.975	17.096
70%	13.672	14.029	14.209	14.439	14.705	14.937	15.107	15.111	15.290
60%	13.295	13.393	13.478	13.566	13.597	13.552	13.629	13.790	13.772
50%	12.955	12.873	12.847	12.820	12.666	12.591	12.534	12.514	12.547
40%	12.623	12.388	12.178	12.014	11.722	11.471	11.290	11.362	11.311
30%	12.262	11.851	11.535	11.227	10.909	10.576	10.364	10.166	10.073
20%	11.828	11.264	10.837	10.423	9.991	9.555	9.288	8.965	8.786
10%	11.298	10.483	9.871	9.330	8.746	8.297	7.860	7.558	7.143
5%	10.840	9.895	9.120	8.436	7.849	7.358	6.750	6.380	5.967
1%	10.030	8.791	7.993	7.279	6.363	5.659	5.069	4.629	4.382

the $S_{|z|}$ and $A_{|z|}$ statistics, while Table 2 reports those for the S_C and A_C statistics considering the 13 autocontours $C = [0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99]$. The percentiles for the S_L and A_L tests considering five lags are reported in Table 1 of the supplementary material.

Since a model can be considered misspecified whenever it does not take into account breaks in the data, the proposed stability tests can be also viewed as statistics that assess the specification and time stability of the model jointly. In practice, the tests will be most useful when models that are estimated using R observations need to be validated over a new set of observations (in our case, over the prediction sample of P observations), and as part of doing so, we also assess whether the model is stable over different time periods. The following sections on Monte Carlo simulations and the evaluation of the Phillips curve show how to use the stability tests.

3. Monte Carlo simulations

We perform extensive Monte Carlo simulations to assess the finite sample properties (size and power) of the proposed statistics. For $\sqrt{R}$-consistent estimators of the parameters of the model, i.e., $(\widehat{\theta} - \theta_0) = O_p(R^{-1/2})$ with a well-defined asymptotic distribution, and under assumption A1, the parameter uncertainty is asymptotically negligible. In this case, the critical values tabulated in Tables 1 and 2 can be used directly. Whenever the condition $P/R \to 0$ is violated, we need to adjust the asymptotic variance of the tests or else the size of the tests will be distorted. As was mentioned in the previous section, the adjustments to the variance will be difficult to calculate analytically because they are model-dependent. For practical purposes, we propose to implement a fully parametric bootstrap procedure because the density model is fully known under the null hypothesis of correct dynamic specification and correct functional form of the conditional density, i.e., stability of the density model over the prediction sample. The bootstrap procedure entails either bootstrapping the asymptotic variance and using the tabulated critical values or bootstrapping the distribution of the proposed Sup- and Ave- tests directly to obtain their critical values. Appendix B explains how we implement the parametric bootstrap in order to obtain bootstrapped critical values. The simulations that follow also explore the effects of different values of the ratio P/R on the size of the tests. We consider fixed, rolling, and recursive forecasting schemes.

3.1. Size of the tests

Under the null hypothesis of a stable density model, we consider the following data generating process: $y_t = \alpha_1+\beta_1 y_{t-1}+\beta_2 x_{t-1}+\sigma\epsilon_t$, where $x_t = \phi_1+\phi_2 x_{t-1}+\nu_t$, and ε_t and $\nu_t \sim N(0, 1)$, $\phi_1 = 1.38$, $\phi_2 = 0.77$, $\alpha_1 = 1.5$, $\beta_1 = 0.5$, $\beta_2 = 0.6$, $\sigma = 1$. We consider sample sizes of $T = 150$

Table 3
Sizes of the statistics: $T = 120, R = 96, P = 24, P/R = 1/4, m = 1/3$ (nominal size 5%).

I. Small sample (T), small subsample window (m), and $P/R = 1/4$.

$T = 120, R = 96, P/R = 1/4$	$S_{\|z\|}^{1,1}$	$S_{\|z\|}^{1,2}$	$S_{\|z\|}^{1,3}$	$S_{\|z\|}^{1,4}$	$S_{\|z\|}^{1,5}$	$S_{\|z\|}^{1,6}$	$S_{\|z\|}^{1,7}$	$S_{\|z\|}^{1,8}$	$S_{\|z\|}^{1,9}$	$S_{\|z\|}^{1,10}$	$S_{\|z\|}^{1,11}$	$S_{\|z\|}^{1,12}$	$S_{\|z\|}^{1,13}$
Fixed	0.015	0.023	0.024	0.032	0.038	0.034	0.032	0.033	0.037	0.031	0.025	0.023	0.016
Rolling	0.012	0.019	0.016	0.028	0.031	0.035	0.033	0.026	0.034	0.032	0.023	0.018	0.013
Recursive	0.014	0.021	0.017	0.030	0.029	0.032	0.031	0.028	0.030	0.030	0.015	0.016	0.015
	$A_{\|z\|}^{1,1}$	$A_{\|z\|}^{1,2}$	$A_{\|z\|}^{1,3}$	$A_{\|z\|}^{1,4}$	$A_{\|z\|}^{1,5}$	$A_{\|z\|}^{1,6}$	$A_{\|z\|}^{1,7}$	$A_{\|z\|}^{1,8}$	$A_{\|z\|}^{1,9}$	$A_{\|z\|}^{1,10}$	$A_{\|z\|}^{1,11}$	$A_{\|z\|}^{1,12}$	$A_{\|z\|}^{1,13}$
Fixed	0.029	0.032	0.040	0.047	0.043	0.043	0.044	0.044	0.046	0.052	0.029	0.032	0.021
Rolling	0.023	0.025	0.025	0.035	0.045	0.045	0.040	0.045	0.045	0.041	0.027	0.025	0.021
Recursive	0.021	0.028	0.024	0.031	0.043	0.046	0.033	0.042	0.045	0.042	0.027	0.027	0.023
	$S_C^{1,13}$	$A_C^{1,13}$	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_L^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$			
Fixed	0.038	0.032	0.030	0.034	0.032	0.030	0.045	0.041	0.042	0.042			
Rolling	0.032	0.033	0.032	0.037	0.035	0.036	0.044	0.042	0.051	0.046			
Recursive	0.029	0.031	0.035	0.033	0.034	0.032	0.045	0.044	0.039	0.042			
	$S_{\|z\|}^{1,7}$	$S_{\|z\|}^{2,7}$	$S_{\|z\|}^{3,7}$	$S_{\|z\|}^{4,7}$	$S_{\|z\|}^{5,7}$	$A_{\|z\|}^{1,7}$	$A_{\|z\|}^{2,7}$	$A_{\|z\|}^{3,7}$	$A_{\|z\|}^{4,7}$	$A_{\|z\|}^{5,7}$			
Fixed	0.032	0.033	0.037	0.032	0.034	0.044	0.043	0.044	0.042	0.039			
Rolling	0.033	0.032	0.033	0.034	0.031	0.041	0.042	0.040	0.039	0.038			
Recursive	0.031	0.030	0.029	0.032	0.029	0.033	0.032	0.031	0.034	0.031			

Notation: 13 autocontours, $C = [0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99]$.
In $S_{|z|}^{k,\alpha_i}$ and $A_{|z|}^{k,\alpha_i}$, the lag k and autocontour α_i are fixed.
In $S_C^{k,\alpha}$ and $A_C^{k,\alpha}$, k is fixed and α is the total number of autocontours considered.
In S_L^{k,α_i} and A_L^{k,α_i}, up to k lags are considered and α_i is a fixed autocontour.

(with evaluation sample $P = 60$), $T = 375$ (with $P = 150$), and $T = 750$ (with $P = 300$) observations, and for each sample size, we consider the proportions $m = r/P$ equal to $m = 1/3$, $m = 1/2$ and $m = 2/3$. We keep the ratio P/R constant and equal to 2/3. A total of nine experiments are run. We work with 13 autocontour coverage levels, $C = [0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99]$. The number of Monte Carlo replications is 1000, the number of bootstrap samples is 500, and the nominal size is 5%. The overall size of the tests is very good in the nine experiments considered. There are no substantial differences among the fixed, rolling, and recursive estimation schemes. We find some small size distortions (undersized) when the sample is small ($T = 150$) and the autocontours levels are extreme (1% and 99%), but the distortion disappears as the sample size increases. For the individual tests $S_{|z|}^{k,\alpha_i}$ and $A_{|z|}^{k,\alpha_i}$ (with k and α_i fixed), the Ave-test tends to have a better size than the Sup-test. The tests S_C, A_C, S_L, and A_L have very good sizes even in small samples. For details, see Tables 2 to 10 in the supplementary material.

For the generating process explained in the previous paragraphs, with a small estimation sample $R = 96$ and a small subsample window $m = 1/3$, we have considered three increasing values for the ratio P/R, namely 1/4 ($T = 120$), 1/2 ($T = 144$), and 3/4 ($T = 168$). The size results for these three cases are presented in Tables 3–5. For the case of $P/R = 1/4$, we observe that the individual tests $S_{|z|}^{k,\alpha_i}$ are more undersized than the tests $A_{|z|}^{k,\alpha_i}$, for which the average size is about 4%. The portmanteau tests S_C, A_C, S_L, and A_L have sizes of between 3% and 5%. Note that we have only a very small prediction sample $P = 24$ in this case, and the subsample window includes only eight observations. There is an overall improvement in size for all tests for the case $P/R = 1/2$, while most tests have sizes around the nominal size of 5% for the case $P/R = 3/4$. It seems that the small distortions that we observe are due to the small number of observations in the subsample window. As we increase the ratio P/R, the parametric bootstrap is properly taking care of the effect of parameter uncertainty in the variance of the tests, and delivering tests with the correct sizes.

3.2. Power of the tests

We assess the power of the tests by generating data from four different processes, all of which contain a break point. The model that we maintain under the null hypothesis is the same as that considered in the study of the size properties: $y_t = \alpha_1 + \beta_1 y_{t-1} + \beta_2 x_{t-1} + \sigma\epsilon_t$, with $x_t = \phi_1 + \phi_2 x_{t-1} + \nu_t$, ν_t and $\varepsilon_t \sim N(0, 1)$. The total sample size (T) is 650, $R = 350$, $P = 300$, and $m = 1/3$. The break point happens at $R + \tau P$, where $\tau = 1/3$, 1/2, and 2/3. In the following experiments, the number of Monte Carlo replications is 1000 and the number of bootstrapped samples is 500. We maintain a nominal test size of 5%.

The four data generating mechanisms are as follows:

DGP1: Break in the intercept of $y_t = \alpha_t + \beta_1 y_{t-1} + \beta_2 x_{t-1} + \sigma\epsilon_t$, $\epsilon_t \sim N(0, 1)$:

$$\alpha_t = \begin{cases} \alpha_1 = 1.5 & \text{if } t < \text{break} \\ \alpha_2 = 2 & \text{otherwise,} \end{cases}$$

with $\beta_1 = 0.5$, $\beta_2 = 0.6$, $\sigma = 1$.

Table 4
Sizes of the statistics: $T = 144$, $R = 96$, $P = 48$, $P/R = 1/2$, $m = 1/3$ (nominal size 5%).

II. Small sample (T), small subsample window (m), and $P/R = 1/2$.													
$T = 144, R = 96, P/R = 1/2$	$S_{\|z\|}^{1,1}$	$S_{\|z\|}^{1,2}$	$S_{\|z\|}^{1,3}$	$S_{\|z\|}^{1,4}$	$S_{\|z\|}^{1,5}$	$S_{\|z\|}^{1,6}$	$S_{\|z\|}^{1,7}$	$S_{\|z\|}^{1,8}$	$S_{\|z\|}^{1,9}$	$S_{\|z\|}^{1,10}$	$S_{\|z\|}^{1,11}$	$S_{\|z\|}^{1,12}$	$S_{\|z\|}^{1,13}$
Fixed	0.022	0.029	0.032	0.037	0.041	0.042	0.034	0.039	0.040	0.044	0.027	0.026	0.023
Rolling	0.019	0.026	0.027	0.035	0.039	0.041	0.039	0.032	0.041	0.046	0.038	0.032	0.020
Recursive	0.021	0.029	0.026	0.032	0.034	0.035	0.035	0.033	0.042	0.043	0.034	0.021	0.023
	$A_{\|z\|}^{1,1}$	$A_{\|z\|}^{1,2}$	$A_{\|z\|}^{1,3}$	$A_{\|z\|}^{1,4}$	$A_{\|z\|}^{1,5}$	$A_{\|z\|}^{1,6}$	$A_{\|z\|}^{1,7}$	$A_{\|z\|}^{1,8}$	$A_{\|z\|}^{1,9}$	$A_{\|z\|}^{1,10}$	$A_{\|z\|}^{1,11}$	$A_{\|z\|}^{1,12}$	$A_{\|z\|}^{1,13}$
Fixed	0.034	0.035	0.040	0.043	0.048	0.056	0.043	0.044	0.040	0.047	0.036	0.033	0.027
Rolling	0.032	0.034	0.033	0.045	0.043	0.041	0.041	0.046	0.043	0.044	0.033	0.032	0.028
Recursive	0.031	0.032	0.030	0.046	0.042	0.042	0.040	0.045	0.047	0.045	0.034	0.030	0.032
	$S_C^{1,13}$	$A_C^{1,13}$	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_L^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$			
Fixed	0.037	0.044	0.037	0.036	0.038	0.038	0.043	0.042	0.046	0.046			
Rolling	0.034	0.042	0.039	0.037	0.037	0.036	0.041	0.045	0.042	0.044			
Recursive	0.031	0.039	0.032	0.033	0.034	0.034	0.042	0.043	0.044	0.042			
	$S_{\|z\|}^{1,7}$	$S_{\|z\|}^{2,7}$	$S_{\|z\|}^{3,7}$	$S_{\|z\|}^{4,7}$	$S_{\|z\|}^{5,7}$	$A_{\|z\|}^{1,7}$	$A_{\|z\|}^{2,7}$	$A_{\|z\|}^{3,7}$	$A_{\|z\|}^{4,7}$	$A_{\|z\|}^{5,7}$			
Fixed	0.035	0.040	0.038	0.039	0.038	0.043	0.050	0.044	0.049	0.043			
Rolling	0.039	0.036	0.037	0.037	0.039	0.041	0.048	0.041	0.039	0.040			
Recursive	0.035	0.039	0.033	0.038	0.036	0.040	0.041	0.039	0.042	0.039			

Notation: 13 autocontours, $C = [0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99]$.
In $S_{|z|}^{k,\alpha_i}$ and $A_{|z|}^{k,\alpha_i}$, the lag k and autocontour α_i are fixed.
In $S_C^{k,\alpha}$ and $A_C^{k,\alpha}$, k is fixed and α is the total number of autocontours considered.
In S_L^{k,α_i} and A_L^{k,α_i}, up to k lags are considered and α_i is a fixed autocontour.

Table 5
Sizes of the statistics: $T = 168$, $R = 96$, $P = 72$, $P/R = 3/4$, $m = 1/3$ (nominal size 5%).

III. Small sample (T), small subsample window (m), and $P/R = 3/4$.													
$T = 168, R = 96, P/R = 3/4$	$S_{\|z\|}^{1,1}$	$S_{\|z\|}^{1,2}$	$S_{\|z\|}^{1,3}$	$S_{\|z\|}^{1,4}$	$S_{\|z\|}^{1,5}$	$S_{\|z\|}^{1,6}$	$S_{\|z\|}^{1,7}$	$S_{\|z\|}^{1,8}$	$S_{\|z\|}^{1,9}$	$S_{\|z\|}^{1,10}$	$S_{\|z\|}^{1,11}$	$S_{\|z\|}^{1,12}$	$S_{\|z\|}^{1,13}$
Fixed	0.032	0.041	0.040	0.042	0.043	0.042	0.041	0.046	0.046	0.046	0.043	0.032	0.039
Rolling	0.031	0.040	0.044	0.040	0.040	0.046	0.046	0.049	0.051	0.046	0.051	0.031	0.038
Recursive	0.029	0.042	0.043	0.053	0.052	0.049	0.043	0.043	0.042	0.042	0.039	0.036	0.035
	$A_{\|z\|}^{1,1}$	$A_{\|z\|}^{1,2}$	$A_{\|z\|}^{1,3}$	$A_{\|z\|}^{1,4}$	$A_{\|z\|}^{1,5}$	$A_{\|z\|}^{1,6}$	$A_{\|z\|}^{1,7}$	$A_{\|z\|}^{1,8}$	$A_{\|z\|}^{1,9}$	$A_{\|z\|}^{1,10}$	$A_{\|z\|}^{1,11}$	$A_{\|z\|}^{1,12}$	$A_{\|z\|}^{1,13}$
Fixed	0.042	0.046	0.047	0.043	0.042	0.043	0.041	0.480	0.052	0.045	0.041	0.038	0.038
Rolling	0.037	0.039	0.040	0.045	0.045	0.047	0.046	0.047	0.051	0.042	0.043	0.037	0.039
Recursive	0.038	0.042	0.037	0.048	0.049	0.047	0.042	0.045	0.049	0.440	0.039	0.036	0.038
	$S_C^{1,13}$	$A_C^{1,13}$	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_L^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$			
Fixed	0.040	0.044	0.043	0.044	0.043	0.045	0.046	0.046	0.045	0.047			
Rolling	0.040	0.052	0.043	0.050	0.044	0.049	0.049	0.044	0.045	0.450			
Recursive	0.042	0.052	0.046	0.043	0.042	0.040	0.043	0.045	0.044	0.045			
	$S_{\|z\|}^{1,7}$	$S_{\|z\|}^{2,7}$	$S_{\|z\|}^{3,7}$	$S_{\|z\|}^{4,7}$	$S_{\|z\|}^{5,7}$	$A_{\|z\|}^{1,7}$	$A_{\|z\|}^{2,7}$	$A_{\|z\|}^{3,7}$	$A_{\|z\|}^{4,7}$	$A_{\|z\|}^{5,7}$			
Fixed	0.041	0.047	0.046	0.044	0.045	0.041	0.041	0.045	0.043	0.045			
Rolling	0.046	0.040	0.042	0.043	0.044	0.046	0.048	0.046	0.045	0.043			
Recursive	0.043	0.041	0.042	0.046	0.044	0.042	0.043	0.045	0.044	0.042			

Notation: 13 autocontours, $C = [0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99]$.
In $S_{|z|}^{k,\alpha_i}$ and $A_{|z|}^{k,\alpha_i}$, the lag k and autocontour α_i are fixed.
In $S_C^{k,\alpha}$ and $A_C^{k,\alpha}$, k is fixed and α is the total number of autocontours considered.
In S_L^{k,α_i} and A_L^{k,α_i}, up to k lags are considered and α_i is a fixed autocontour.

DGP2: Break in the variance of $y_t = \alpha + \beta_1 y_{t-1} + \beta_2 x_{t-1} + \sigma_t \epsilon_t$, $\epsilon_t \sim N(0, 1)$:

$$\sigma_t = \begin{cases} \sigma_1 = 1.5 & \text{if } t < \text{break} \\ \sigma_2 = 1.8 & \text{otherwise,} \end{cases}$$

with $\alpha = 1.5$, $\beta_1 = 0.5$, $\beta_2 = 0.6$.

DGP3: Breaks in the slope coefficients of $y_t = \alpha + \beta_{1,t} y_{t-1} + \beta_{2,t} x_{t-1} + \sigma \epsilon_t$, $\epsilon_t \sim N(0, 1)$:

$$\beta_{1,t} = \begin{cases} \beta_{1,1} = 0.5 & \text{if } t < \text{break} \\ \beta_{1,2} = 0.3 & \text{otherwise} \end{cases}$$

Table 6
Power for DGP1 under the fixed and rolling schemes.

Fixed scheme	$S_{\|z\|}^{l,1}$	$S_{\|z\|}^{l,2}$	$S_{\|z\|}^{l,3}$	$S_{\|z\|}^{l,4}$	$S_{\|z\|}^{l,5}$	$S_{\|z\|}^{l,6}$	$S_{\|z\|}^{l,7}$	$S_{\|z\|}^{l,8}$	$S_{\|z\|}^{l,9}$	$S_{\|z\|}^{l,10}$	$S_{\|z\|}^{l,11}$	$S_{\|z\|}^{l,12}$	$S_{\|z\|}^{l,13}$
$l=1$	0.14	0.450	0.765	0.956	0.979	0.987	0.995	0.992	0.987	0.976	0.938	0.867	0.589
$l=2$	0.14	0.464	0.789	0.955	0.981	0.990	0.993	0.993	0.988	0.974	0.943	0.864	0.593
$l=3$	0.14	0.419	0.792	0.965	0.985	0.997	0.997	0.994	0.985	0.971	0.937	0.871	0.594
$l=4$	0.11	0.439	0.797	0.953	0.986	0.993	0.994	0.994	0.990	0.979	0.938	0.866	0.595
$l=5$	0.12	0.453	0.796	0.952	0.984	0.992	0.994	0.993	0.989	0.971	0.938	0.867	0.598
	$A_{\|z\|}^{l,1}$	$A_{\|z\|}^{l,2}$	$A_{\|z\|}^{l,3}$	$A_{\|z\|}^{l,4}$	$A_{\|z\|}^{l,5}$	$A_{\|z\|}^{l,6}$	$A_{\|z\|}^{l,7}$	$A_{\|z\|}^{l,8}$	$A_{\|z\|}^{l,9}$	$A_{\|z\|}^{l,10}$	$A_{\|z\|}^{l,11}$	$A_{\|z\|}^{l,12}$	$A_{\|z\|}^{l,13}$
$l=1$	0.08	0.244	0.698	0.898	0.948	0.969	0.981	0.978	0.968	0.945	0.904	0.812	0.588
$l=2$	0.09	0.277	0.741	0.896	0.952	0.965	0.979	0.983	0.972	0.947	0.898	0.815	0.585
$l=3$	0.07	0.276	0.720	0.893	0.941	0.969	0.980	0.984	0.967	0.943	0.906	0.815	0.588
$l=4$	0.07	0.319	0.702	0.887	0.957	0.971	0.979	0.981	0.971	0.944	0.904	0.817	0.586
$l=5$	0.08	0.339	0.722	0.906	0.951	0.968	0.985	0.976	0.972	0.944	0.903	0.816	0.584
	$S_C^{l,13}$	$A_C^{l,13}$											
$l=1$	0.961	0.921											
$l=2$	0.965	0.924											
$l=3$	0.967	0.923											
$l=4$	0.963	0.921											
$l=5$	0.965	0.921											
	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_L^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$					
$C=7$	0.989	0.985	0.976	0.964	0.980	0.969	0.956	0.942					
Rolling scheme	$S_{\|z\|}^{l,1}$	$S_{\|z\|}^{l,2}$	$S_{\|z\|}^{l,3}$	$S_{\|z\|}^{l,4}$	$S_{\|z\|}^{l,5}$	$S_{\|z\|}^{l,6}$	$S_{\|z\|}^{l,7}$	$S_{\|z\|}^{l,8}$	$S_{\|z\|}^{l,9}$	$S_{\|z\|}^{l,10}$	$S_{\|z\|}^{l,11}$	$S_{\|z\|}^{l,12}$	$S_{\|z\|}^{l,13}$
$l=1$	0.18	0.12	0.308	0.416	0.459	0.394	0.420	0.264	0.210	0.080	0.119	0.092	0.080
$l=2$	0.11	0.11	0.295	0.324	0.323	0.431	0.433	0.333	0.193	0.084	0.123	0.093	0.080
$l=3$	0.11	0.11	0.258	0.402	0.386	0.390	0.472	0.325	0.207	0.096	0.123	0.094	0.090
$l=4$	0.13	0.11	0.289	0.439	0.410	0.413	0.378	0.325	0.173	0.093	0.121	0.092	0.082
$l=5$	0.16	0.11	0.262	0.445	0.418	0.423	0.438	0.318	0.171	0.091	0.121	0.092	0.083
	$A_{\|z\|}^{l,1}$	$A_{\|z\|}^{l,2}$	$A_{\|z\|}^{l,3}$	$A_{\|z\|}^{l,4}$	$A_{\|z\|}^{l,5}$	$A_{\|z\|}^{l,6}$	$A_{\|z\|}^{l,7}$	$A_{\|z\|}^{l,8}$	$A_{\|z\|}^{l,9}$	$A_{\|z\|}^{l,10}$	$A_{\|z\|}^{l,11}$	$A_{\|z\|}^{l,12}$	$A_{\|z\|}^{l,13}$
$l=1$	0.25	0.318	0.559	0.783	0.738	0.710	0.534	0.634	0.418	0.18	0.21	0.102	0.091
$l=2$	0.22	0.292	0.527	0.733	0.71	0.714	0.56	0.621	0.438	0.17	0.216	0.100	0.092
$l=3$	0.24	0.275	0.521	0.768	0.756	0.744	0.499	0.638	0.442	0.155	0.224	0.103	0.092
$l=4$	0.25	0.28	0.519	0.806	0.775	0.759	0.611	0.638	0.421	0.133	0.21	0.101	0.092
$l=5$	0.25	0.305	0.516	0.797	0.781	0.769	0.533	0.626	0.392	0.121	0.24	0.101	0.093
	$S_C^{l,13,}$	$A_C^{l,13}$											
$l=1$	0.638	0.749											
$l=2$	0.521	0.733											
$l=3$	0.619	0.801											
$l=4$	0.641	0.822											
$l=5$	0.637	0.803											
	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_L^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$					
$C=7$	0.414	0.419	0.408	0.432	0.563	0.59	0.578	0.532					

Notes: 13 autocontours, $C = [0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99]$.
$S_{|z|}^{l,7}, A_{|z|}^{l,7}$ for $l = 1, 2, \ldots, 5$; 7 refers to the 50% autocontour.
$S_L^{l,7}, A_L^{l,7}$ stacking lags up to $l = 2, \ldots .5$ and considering the 50% autocontour.
$S_C^{l,13}$ and $A_C^{l,13}$ stacking all 13 autocontours for one lag $l = 1, 2, 3, 4, 5$.
1000 Monte Carlo replications and 500 bootstrap samples.
$T = 650$, $R = 350$, $P = 300$, $m = 1/3$, and a break point at $R + \tau P$ for $\tau = 1/3$.

$$\beta_{2,t} = \begin{cases} \beta_{2,1} = 0.6 & \text{if } t < \text{ break} \\ \beta_{2,2} = 0.4 & \text{otherwise,} \end{cases}$$

with $\alpha = 1.5, \sigma = 1$.

DGP4: Breaks in the intercept, variance and slope coefficients of $y_t = \alpha_t + \beta_{1,t} y_{t-1} + \beta_{2,t} x_{t-1} + \sigma_t \epsilon_t$, $\epsilon_t \sim N(0, 1)$:

$$\alpha_t = \begin{cases} \alpha_1 = 1.5 & \text{if } t < \text{ break} \\ \alpha_2 = 2 & \text{otherwise} \end{cases}$$

$$\sigma_t = \begin{cases} \sigma_1 = 1.5 & \text{if } t < \text{ break} \\ \sigma_2 = 1.8 & \text{otherwise} \end{cases}$$

$$\beta_{1,t} = \begin{cases} \beta_{1,1} = 0.5 & \text{if } t < \text{ break} \\ \beta_{1,2} = 0.3 & \text{otherwise} \end{cases}$$

$$\beta_{2,t} = \begin{cases} \beta_{2,1} = 0.6 & \text{if } t < \text{ break} \\ \beta_{2,2} = 0.4 & \text{otherwise.} \end{cases}$$

Note that the breaks considered are not very extreme. We perform all the simulations under fixed, rolling, and recursive estimation schemes, and report the power results for the fixed and rolling ($\tau = 1/3$) schemes for DGPs 1 and 2 in Tables 6 and 7. The power results for DGPs 3 and 4 under the fixed, rolling, and recursive schemes are reported in Tables 11 to 14 of the supplementary material.

Table 7
Power for DGP2 under the fixed and rolling schemes.

Fixed scheme	$S^{l,1}_{\|z\|}$	$S^{l,2}_{\|z\|}$	$S^{l,3}_{\|z\|}$	$S^{l,4}_{\|z\|}$	$S^{l,5}_{\|z\|}$	$S^{l,6}_{\|z\|}$	$S^{l,7}_{\|z\|}$	$S^{l,8}_{\|z\|}$	$S^{l,9}_{\|z\|}$	$S^{l,10}_{\|z\|}$	$S^{l,11}_{\|z\|}$	$S^{l,12}_{\|z\|}$	$S^{l,13}_{\|z\|}$
$l = 1$	0.292	0.271	0.191	0.084	0.073	0.109	0.281	0.278	0.391	0.512	0.638	0.638	0.518
$l = 2$	0.260	0.239	0.171	0.079	0.075	0.106	0.282	0.292	0.418	0.518	0.637	0.639	0.520
$l = 3$	0.285	0.265	0.175	0.089	0.075	0.120	0.283	0.290	0.404	0.530	0.635	0.639	0.523
$l = 4$	0.311	0.260	0.162	0.088	0.063	0.096	0.282	0.284	0.399	0.512	0.624	0.636	0.526
$l = 5$	0.273	0.253	0.158	0.068	0.068	0.108	0.282	0.288	0.408	0.539	0.645	0.638	0.531
	$A^{l,1}_{\|z\|}$	$A^{l,2}_{\|z\|}$	$A^{l,3}_{\|z\|}$	$A^{l,4}_{\|z\|}$	$A^{l,5}_{\|z\|}$	$A^{l,6}_{\|z\|}$	$A^{l,7}_{\|z\|}$	$A^{l,8}_{\|z\|}$	$A^{l,9}_{\|z\|}$	$A^{l,10}_{\|z\|}$	$A^{l,11}_{\|z\|}$	$A^{l,12}_{\|z\|}$	$A^{l,13}_{\|z\|}$
$l = 1$	0.345	0.242	0.162	0.090	0.078	0.127	0.275	0.271	0.361	0.472	0.561	0.604	0.527
$l = 2$	0.327	0.225	0.160	0.081	0.077	0.120	0.269	0.277	0.377	0.467	0.560	0.606	0.522
$l = 3$	0.347	0.239	0.167	0.089	0.087	0.123	0.272	0.286	0.374	0.468	0.558	0.600	0.527
$l = 4$	0.368	0.246	0.159	0.090	0.069	0.111	0.272	0.277	0.345	0.455	0.550	0.592	0.523
$l = 5$	0.317	0.222	0.146	0.086	0.072	0.120	0.273	0.288	0.370	0.469	0.564	0.600	0.523
	$S^{l,13}_{C}$	$A^{l,13}_{C}$											
$l = 1$	0.659	0.595											
$l = 2$	0.651	0.604											
$l = 3$	0.666	0.625											
$l = 4$	0.644	0.625											
$l = 5$	0.652	0.611											
	$S^{2,7}_{L}$	$S^{3,7}_{L}$	$S^{4,7}_{L}$	$S^{5,7}_{L}$	$A^{2,7}_{L}$	$A^{3,7}_{L}$	$A^{4,7}_{L}$	$A^{5,7}_{L}$					
$C = 7$	0.286	0.285	0.281	0.285	0.271	0.273	0.276	0.275					
Rolling scheme	$S^{l,1}_{\|z\|}$	$S^{l,2}_{\|z\|}$	$S^{l,3}_{\|z\|}$	$S^{l,4}_{\|z\|}$	$S^{l,5}_{\|z\|}$	$S^{l,6}_{\|z\|}$	$S^{l,7}_{\|z\|}$	$S^{l,8}_{\|z\|}$	$S^{l,9}_{\|z\|}$	$S^{l,10}_{\|z\|}$	$S^{l,11}_{\|z\|}$	$S^{l,12}_{\|z\|}$	$S^{l,13}_{\|z\|}$
$l = 1$	0.116	0.104	0.088	0.063	0.050	0.056	0.112	0.110	0.171	0.341	0.302	0.264	0.261
$l = 2$	0.114	0.090	0.086	0.063	0.054	0.051	0.118	0.113	0.172	0.318	0.296	0.264	0.262
$l = 3$	0.107	0.094	0.076	0.065	0.060	0.053	0.108	0.126	0.189	0.335	0.280	0.273	0.266
$l = 4$	0.113	0.104	0.072	0.054	0.058	0.049	0.175	0.114	0.167	0.329	0.287	0.266	0.267
$l = 5$	0.109	0.093	0.080	0.053	0.050	0.053	0.160	0.121	0.184	0.337	0.274	0.265	0.267
	$A^{l,1}_{\|z\|}$	$A^{l,2}_{\|z\|}$	$A^{l,3}_{\|z\|}$	$A^{l,4}_{\|z\|}$	$A^{l,5}_{\|z\|}$	$A^{l,6}_{\|z\|}$	$A^{l,7}_{\|z\|}$	$A^{l,8}_{\|z\|}$	$A^{l,9}_{\|z\|}$	$A^{l,10}_{\|z\|}$	$A^{l,11}_{\|z\|}$	$A^{l,12}_{\|z\|}$	$A^{l,13}_{\|z\|}$
$l = 1$	0.179	0.129	0.099	0.078	0.066	0.075	0.190	0.151	0.204	0.360	0.344	0.350	0.378
$l = 2$	0.169	0.112	0.082	0.075	0.066	0.076	0.180	0.145	0.205	0.386	0.345	0.343	0.379
$l = 3$	0.175	0.139	0.092	0.070	0.067	0.077	0.196	0.159	0.228	0.378	0.352	0.348	0.378
$l = 4$	0.183	0.143	0.083	0.063	0.066	0.077	0.192	0.144	0.201	0.358	0.342	0.352	0.375
$l = 5$	0.159	0.129	0.090	0.063	0.066	0.065	0.191	0.151	0.208	0.371	0.347	0.350	0.377
	$S^{l,13}_{C}$	$A^{l,13}_{C}$											
$l = 1$	0.409	0.414											
$l = 2$	0.363	0.394											
$l = 3$	0.381	0.427											
$l = 4$	0.398	0.431											
$l = 5$	0.306	0.318											
	$S^{2,7}_{L}$	$S^{3,7}_{L}$	$S^{4,7}_{L}$	$S^{5,7}_{L}$	$A^{2,7}_{L}$	$A^{3,7}_{L}$	$A^{4,7}_{L}$	$A^{5,7}_{L}$					
$C = 7$	0.168	0.115	0.113	0.112	0.129	0.189	0.191	0.182					

Notes: 13 autocontours, $C = [0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99]$.
$S^{l,7}_{|z|}, A^{l,7}_{|z|}$ for $l = 1, 2, \ldots, 5$; 7 refers to the 50% autocontour.
$S^{l,7}_{L}, A^{l,7}_{L}$ stacking lags up to $l = 2, \ldots .5$ and considering the 50% autocontour.
$S^{l,13}_{C}$ and $A^{l,13}_{C}$ stacking all 13 autocontours for one lag $l = 1, 2, 3, 4, 5$.
1000 Monte Carlo replications and 500 bootstrap samples.
$T = 650$, $R = 350$, $P = 300$, $m = 1/3$, and a break point at $R + \tau P$ for $\tau = 1/3$.

Under the fixed scheme, the tests are most powerful (power of about 90%) for detecting breaks in the intercept and slope coefficients (DGP1, DGP3, and DGP4). The Ave- and Sup-tests perform similarly, for either the individual hypotheses ($S_{|z|}$ and $A_{|z|}$) or the joint hypothesis (S_C, A_C, S_L and A_L). The power drops when testing for breaks in the variance (DGP2), with values of 10%–60% for the single hypothesis tests and 60% for the joint hypothesis tests. As expected, the tests are less powerful overall under the rolling scheme, because rolling the estimation sample means that the model slowly adjusts to the new parameter values. Nevertheless, S_C, A_C, S_L and A_L are still very powerful (around 50%–80%) for DGP1, DGP3, and DGP4, and around 40% for DGP2. The Ave-test is more powerful than the Sup-test for all DGPs. Under the recursive scheme, the tests perform slightly worse than under the fixed scheme, but slightly better than under the rolling scheme.

We move the location of the break, i.e., $\tau = 1/2$ and $2/3$, and present the power results for DGP3 with a rolling scheme in Table 8. Overall, we find that the power of the

Table 8
Power for DGP3 under the rolling scheme $\tau = 1/2$ and $2/3$.

Rolling scheme $\tau = 1/2$	$S_{\|z\|}^{l,1}$	$S_{\|z\|}^{l,2}$	$S_{\|z\|}^{l,3}$	$S_{\|z\|}^{l,4}$	$S_{\|z\|}^{l,5}$	$S_{\|z\|}^{l,6}$	$S_{\|z\|}^{l,7}$	$S_{\|z\|}^{l,8}$	$S_{\|z\|}^{l,9}$	$S_{\|z\|}^{l,10}$	$S_{\|z\|}^{l,11}$	$S_{\|z\|}^{l,12}$	$S_{\|z\|}^{l,13}$
$l = 1$	0.326	0.322	0.315	0.132	0.198	0.278	0.352	0.425	0.368	0.109	0.183	0.081	0.069
$l = 2$	0.305	0.246	0.286	0.304	0.294	0.413	0.442	0.456	0.348	0.113	0.193	0.079	0.072
$l = 3$	0.318	0.322	0.301	0.365	0.348	0.401	0.463	0.427	0.372	0.078	0.191	0.083	0.079
$l = 4$	0.322	0.316	0.325	0.392	0.456	0.487	0.459	0.464	0.301	0.083	0.192	0.087	0.069
$l = 5$	0.341	0.301	0.297	0.387	0.432	0.445	0.476	0.447	0.351	0.069	0.177	0.072	0.069
	$A_{\|z\|}^{l,1}$	$A_{\|z\|}^{l,2}$	$A_{\|z\|}^{l,3}$	$A_{\|z\|}^{l,4}$	$A_{\|z\|}^{l,5}$	$A_{\|z\|}^{l,6}$	$A_{\|z\|}^{l,7}$	$A_{\|z\|}^{l,8}$	$A_{\|z\|}^{l,9}$	$A_{\|z\|}^{l,10}$	$A_{\|z\|}^{l,11}$	$A_{\|z\|}^{l,12}$	$A_{\|z\|}^{l,13}$
$l = 1$	0.143	0.187	0.216	0.342	0.403	0.503	0.631	0.645	0.614	0.314	0.133	0.064	0.052
$l = 2$	0.273	0.613	0.632	0.723	0.765	0.728	0.668	0.727	0.625	0.322	0.124	0.062	0.046
$l = 3$	0.404	0.647	0.687	0.745	0.794	0.716	0.589	0.724	0.653	0.301	0.116	0.067	0.064
$l = 4$	0.382	0.587	0.618	0.742	0.783	0.786	0.586	0.722	0.622	0.246	0.125	0.065	0.053
$l = 5$	0.313	0.578	0.628	0.765	0.784	0.785	0.658	0.729	0.635	0.204	0.119	0.065	0.052
	$S_C^{l,13}$	$A_C^{l,13}$											
$l = 1$	0.513	0.653											
$l = 2$	0.452	0.654											
$l = 3$	0.447	0.649											
$l = 4$	0.501	0.639											
$l = 5$	0.472	0.624											
	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_L^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$					
$C = 7$	0.358	0.442	0.436	0.397	0.694	0.672	0.576	0.542					
Rolling scheme $\tau = 2/3$	$S_{\|z\|}^{l,1}$	$S_{\|z\|}^{l,2}$	$S_{\|z\|}^{l,3}$	$S_{\|z\|}^{l,4}$	$S_{\|z\|}^{l,5}$	$S_{\|z\|}^{l,6}$	$S_{\|z\|}^{l,7}$	$S_{\|z\|}^{l,8}$	$S_{\|z\|}^{l,9}$	$S_{\|z\|}^{l,10}$	$S_{\|z\|}^{l,11}$	$S_{\|z\|}^{l,12}$	$S_{\|z\|}^{l,13}$
$l = 1$	0.221	0.206	0.221	0.106	0.148	0.214	0.268	0.397	0.261	0.096	0.132	0.067	0.056
$l = 2$	0.215	0.194	0.214	0.228	0.226	0.331	0.345	0.402	0.279	0.101	0.152	0.071	0.061
$l = 3$	0.242	0.247	0.205	0.204	0.215	0.347	0.344	0.401	0.244	0.069	0.144	0.068	0.053
$l = 4$	0.212	0.272	0.232	0.224	0.222	0.362	0.353	0.398	0.235	0.073	0.132	0.068	0.058
$l = 5$	0.253	0.265	0.247	0.252	0.253	0.343	0.385	0.408	0.245	0.062	0.147	0.063	0.057
	$A_{\|z\|}^{l,1}$	$A_{\|z\|}^{l,2}$	$A_{\|z\|}^{l,3}$	$A_{\|z\|}^{l,4}$	$A_{\|z\|}^{l,5}$	$A_{\|z\|}^{l,6}$	$A_{\|z\|}^{l,7}$	$A_{\|z\|}^{l,8}$	$A_{\|z\|}^{l,9}$	$A_{\|z\|}^{l,10}$	$A_{\|z\|}^{l,11}$	$A_{\|z\|}^{l,12}$	$A_{\|z\|}^{l,13}$
$l = 1$	0.112	0.154	0.164	0.213	0.361	0.453	0.538	0.597	0.526	0.223	0.103	0.056	0.042
$l = 2$	0.214	0.527	0.487	0.583	0.644	0.682	0.523	0.623	0.535	0.213	0.102	0.054	0.041
$l = 3$	0.253	0.562	0.524	0.572	0.634	0.672	0.474	0.643	0.534	0.223	0.109	0.057	0.042
$l = 4$	0.302	0.475	0.516	0.602	0.657	0.685	0.454	0.638	0.547	0.156	0.105	0.058	0.042
$l = 5$	0.268	0.467	0.537	0.586	0.635	0.674	0.464	0.648	0.518	0.157	0.103	0.057	0.043
	$S_C^{l,13}$	$A_C^{l,13}$											
$l = 1$	0.423	0.544											
$l = 2$	0.413	0.538											
$l = 3$	0.409	0.563											
$l = 4$	0.427	0.557											
$l = 5$	0.418	0.556											
	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_L^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$					
$C = 7$	0.305	0.334	0.354	0.375	0.514	0.523	0.517	0.498					

Notes: 13 autocontours, $C = [0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99]$.
$S_{|z|}^{l,7}, A_{|z|}^{l,7}$ for $l = 1, 2, \dots, 5$; 7 refers to the 50% autocontour.
$S_L^{l,7}, A_L^{l,7}$ stacking lags up to $l = 2, \dots .5$ and considering the 50% autocontour.
$S_C^{l,13}$ and $A_C^{l,13}$ stacking all 13 autocontours for one lag $l = 1, 2, 3, 4, 5$.
1000 Monte Carlo replications and 500 bootstrap samples.
$T = 650$, $R = 350$, $P = 300$, $m = 1/3$, and a break point at $R + \tau P$ for $\tau = 1/2, 2/3$.

tests decreases as the break point moves towards the end of the prediction sample, which is as expected. However, the decrease in power is not very abrupt. For instance, the power of the A_C test moves from about 80% ($\tau = 1/3$) to 65% ($\tau = 1/2$) to 55% ($\tau = 2/3$).

3.3. Comparison with the Rossi & Sekhposyan tests

We compare our tests with those of Rossi & Sekhposyan (R&S) (2013). The R&S (2013) tests and our tests have similar null hypotheses, namely the constancy of the density model over the prediction sample, although their statistics allow for dynamic misspecification. The R&S tests follow the setup of Corradi and Swanson (2006). They are also based on the PITs $\{u_t\}$ of the proposed model and they measure the distance between the empirical cumulative distribution function and that of the uniform distribution, which is a 45 degree line. Under dynamic misspecification, the PITs are still distributed uniformly in [0,1], but are no longer independent. Thus, the R&S tests borrow from Corradi and Swanson, in that the limiting variance of the statistics has to take into account the potential lack of independence. Their κ_P test is a Kolmogorov–Smirnov-type statistic, and their C_P is a Cramer-von Mises-type statistic. Our G-ACR tests are based on the object "autocontour", and measure the independence, denseness and uniform distribution of the PITs over a collection of squares in a two-dimensional space (u_{t-k}, u_t). By construction, our tests exploit the independence properties of the PITs directly. The asymptotic distribution of the R&S tests is based on the statistical properties of an empirical process. Our tests are simpler, in that they rely on the properties of a binary indicator with well-defined moments. When the parameter uncertainty is non-negligible, the critical values of the R&S and G-ACR tests are obtained by simulation.

We follow the experimental design explained in Section 4.2 and Table 2 of the R&S article. Their tests κ_P and C_P are directly comparable with our tests S_C and A_C, stacking the 13 autocontour levels. We consider the same DGPs as in Table 2 of the R&S article: DGP1 considers the case of misspecification in the functional form of the density, DGP2 considers instability in the mean and variance parameters, and DGP3 considers time-varying distortions in the shape of the assumed conditional density. Note that none of these three DGPs contemplates dynamic misspecification. Thus, the comparison between the tests of R&S (2013) and our autocontour-based tests is on fair and equal grounds. We run the experiments with $R = 500$, $P = 100$ and a rolling ratio $m = 0.8$ in the prediction sample. The models are estimated recursively. The critical values for the R&S tests are obtained by simulation (for details, see Table 2 of the R&S article). For our tests, we use 1000 Monte Carlo replications with 500 bootstrap samples.

The results of the comparison are provided in Table 9. When $c = 0$ (DGP1 and DGP2) and $T_1/T = 600$ (DGP3), the respective DGPs coincide with the model under the null hypothesis, meaning that the rejection frequency must be equal to the nominal size (5%) of the tests. The larger the value of c and the lower the ratio T_1/T, the further the DGP model is from the model under the null. All of the tests have good sizes, but the $S_C^{1,13}$ and $A_C^{1,13}$ tests are more powerful across DGPs, especially for the cases of DGP2 ($c = 0.5$ and $c = 1$) and DGP3 ($T_1/T = 580$ and 560).

4. Density forecast evaluation of the Phillips curve

We use the proposed tests to analyze the stability of the Phillips curve. Stock and Watson (1999) found some empirical evidence in favor of using the Phillips curve as a forecasting tool, and showed that the inflation forecasts produced by the Phillips curve were more accurate than those based on simple autoregressive or multivariate models, but they also found parameter instabilities across different subsamples. Rossi and Sekhposyan (2010) showed that the predictive power of the Phillips curve disappeared at around the time of the Great Moderation. Based on scoring rules, Amisano and Giacomini (2007) compared the density forecast accuracy of several models of the Phillips curve, and concluded that the best density forecasts are produced by a Markov-switching model. Since their comparisons are based on the *average* forecasting performance of competing models over time, they cannot address directly the instabilities that are documented widely in the literature. In this section, we consider the models of Amisano and Giacomini (2007), and focus on their absolute density forecast performance in the presence of instabilities. Our starting model is a linear Phillips curve (Stock & Watson, 1999), in which changes in the inflation rate depend on their lags and on lags of the unemployment rate, i.e.,

$$\Delta\pi_t = \alpha_1 + \beta_1\Delta\pi_{t-1} + \beta_2\Delta\pi_{t-2} + \beta_{12}\Delta\pi_{t-12} + \gamma u_{t-1} + \sigma\epsilon_t,$$

where $\pi_t = 100 \times \ln(CPI_t/CPI_{t-12})$; CPI_t is the consumer price index for all urban consumers and all items; u_{t-1} is the civilian unemployment rate; and $\epsilon_t \sim N(0, 1)$. The data are collected from the FRED database; the monthly CPI and unemployment series are both seasonally adjusted. The time series run from 1958M01 to 2012M01 (updated from the 1958M01–2004M07 sample of Amisano & Giacomini). The standard tests on $\Delta\pi_t$ and u_t show that they do not have unit roots. On implementing our tests, we consider the same estimation sample as Amisano and Giacomini, from 1958M01 to 1987M12 (360 observations). The evaluation sample runs from 1988M01 to 2012M01 (289 observations), with subsamples of size $r = 200$.

4.1. Evaluation of the linear Phillips curve

We present the evaluation results for the linear Phillips curve in Table 10 under both fixed and rolling estimation schemes. The recursive case is discussed in the supplementary material.

The portmanteau tests S_C, A_C, S_L and A_L indicate a clear rejection of the linear model. On examining the individual tests $S_{|z|}$ and $A_{|z|}$, the rejection comes from the middle autocontours, between 40% and 70% coverage, and from the 95%–99% levels (large changes in inflation). Fig. 1 plots the t- and C-statistics over the evaluation period in a sequential fashion. The tests break through their corresponding critical values at around the 60th statistic, and the values of the tests keep on increasing, reaching two local maxima, which points to two potential breaks: the first in the 64th statistic for the t-tests (1993M03 to 2009M11) and the 71th statistic for the C-tests (1993M11 to 2010M06), and the second in the 78th statistic, which corresponds to the period 1994M06 to 2011M01. All of these periods include the years 1993–2007 of the so-called Great Moderation and the years following the deep financial crisis of 2008.

4.2. Evaluation of the non-linear Phillips curve

Given the rejection of the linear Phillips curve, we proceed with a flexible specification by assuming that the coefficients in the linear model vary according to a Markov switching mechanism. We consider a two-state Markov switching model, i.e.,

$$\Delta\pi_t = \alpha^{s_t} + \beta_1^{s_t}\Delta\pi_{t-1} + \beta_2^{s_t}\Delta\pi_{t-2} + \beta_{12}^{s_t}\Delta\pi_{t-12} + \gamma^{s_t}u_{t-1} + \sigma^{s_t}\epsilon_t,$$

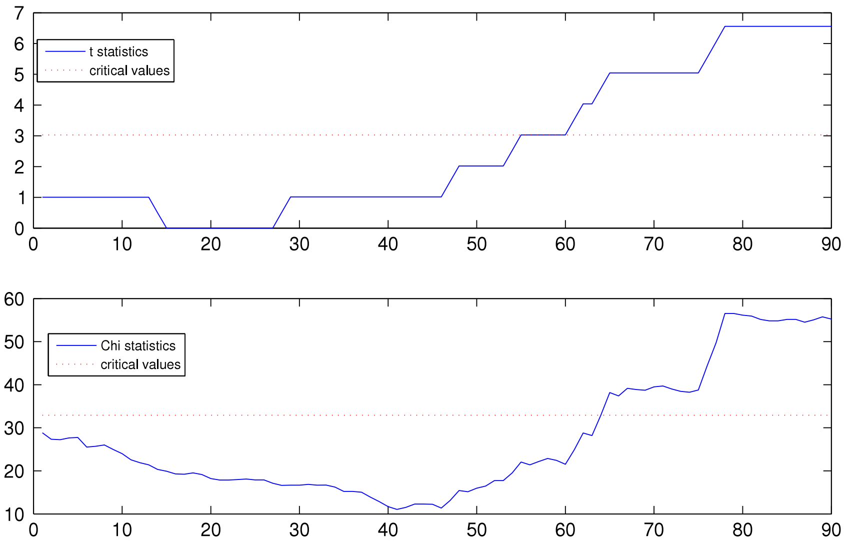

Fig. 1. Plots of t-ratios (99% autocontour) and C-statistics for the linear Phillips curve (fixed scheme). On the x-axis, the number of the evaluation sample in sequential order.

where the unobserved state variable s_t switches between two states, 1 or 2, with transition probabilities $Pr(s_t = j|s_{t-1} = i) = p_{ij}$ for $i, j = 1, 2$; and ϵ_t is assumed to be a standard normal variate. Thus, this model allows for non-Gaussian density forecasts. Since all of the coefficients depend on the state variable (Model 1), the model is extremely flexible and will adapt to any potential breaks or instabilities that may occur over time. We have run our test statistics, and fail to reject the null hypothesis of correct specification. However, we would like to investigate which coefficients are key to understanding where the nonlinear behavior comes from. We consider six additional specifications.

Model 2: $\Delta\pi_t = \alpha + \beta_1^{s_t}\Delta\pi_{t-1} + \beta_2^{s_t}\Delta\pi_{t-2} + \beta_{12}^{s_t}\Delta\pi_{t-12} + \gamma u_{t-1} + \sigma\epsilon_t$ (the intercept and the unemployment coefficient do not depend on s_t).

Model 3: $\Delta\pi_t = \alpha^{s_t} + \beta_1\Delta\pi_{t-1} + \beta_2\Delta\pi_{t-2} + \beta_{12}\Delta\pi_{t-12} + \gamma^{s_t}u_{t-1} + \sigma^{s_t}\epsilon_t$ (the inflation coefficients do not depend on s_t).

Model 4: $\Delta\pi_t = \alpha + \beta_1\Delta\pi_{t-1} + \beta_2\Delta\pi_{t-2} + \beta_{12}\Delta\pi_{t-12} + \gamma u_{t-1} + \sigma^{s_t}\epsilon_t$ (only the standard deviation depends on s_t).

Model 5: $\Delta\pi_t = \alpha^{s_t} + \beta_1\Delta\pi_{t-1} + \beta_2\Delta\pi_{t-2} + \beta_{12}\Delta\pi_{t-12} + \gamma^{s_t}u_{t-1} + \sigma\epsilon_t$ (only the intercept and the unemployment coefficient depend on s_t).

Model 6: $\Delta\pi_t = \alpha^{s_t} + \beta_1\Delta\pi_{t-1} + \beta_2\Delta\pi_{t-2} + \beta_{12}\Delta\pi_{t-12} + \gamma u_{t-1} + \sigma^{s_t}\epsilon_t$ (only the intercept and the standard deviation depend on s_t).

Model 7: $\Delta\pi_t = \alpha + \beta_1\Delta\pi_{t-1} + \beta_2\Delta\pi_{t-2} + \beta_{12}\Delta\pi_{t-12} + \gamma^{s_t}u_{t-1} + \sigma\epsilon_t$ (only the unemployment coefficient depends on s_t).

Of these additional six specifications, we observe outright rejections at the 5% significance level for Model 2 (rejection in the 90%–99% autocontours), Model 4 (rejection in the 95%–99% autocontours), and Model 7 (rejection in the 30%–50% autocontours). The only model that we fail to reject is Model 3, in which we allow the intercept, the unemployment coefficient, and the standard deviation of the error to be state-dependent, while the rest of the parameters are constant. In this case, all statistics (joint and individual) enjoy p-values larger than 5%. Models 5 and 6 are also close contenders. In Model 5, where only the intercept and the unemployment coefficient are state-dependent, we observe a marginal rejection (p-values of around 4%) with the $S_{|z|}$ and $A_{|z|}$ tests for the 1% and 5% autocontours. Likewise, in Model 6, where only the in-

Table 9
Comparison with Rossi & Sekhposyan tests. Empirical rejection frequencies.

	DGP1			
c	κ_P	C_P	$S_C^{1,13}$	$A_C^{1,13}$
0	0.04	0.04	0.05	0.04
0.5	0.04	0.05	0.07	0.07
1	0.13	0.14	0.21	0.20
1.5	0.30	0.34	0.48	0.40
2	0.50	0.60	0.71	0.63
2.5	0.73	0.82	0.87	0.84
	DGP2			
c	κ_P	C_P	$S_C^{1,13}$	$A_C^{1,13}$
0	0.04	0.04	0.05	0.05
0.5	0.16	0.14	0.53	0.50
1	0.66	0.47	0.99	0.97
1.5	0.97	0.92	1.00	1.00
2	1.00	1.00	1.00	1.00
2.5	1.00	1.00	1.00	1.00
	DGP3			
T_1/T	κ_P	C_P	$S_C^{1,13}$	$A_C^{1,13}$
600	0.04	0.04	0.05	0.05
580	0.27	0.29	0.73	0.50
560	0.78	0.78	1.00	0.99
540	0.96	0.97	1.00	1.00

Table 10
Bootstrapped p-values: linear Phillips curve (fixed and rolling schemes).

Fixed scheme	$S_{\lvert z\rvert}^{l,1}$	$S_{\lvert z\rvert}^{l,2}$	$S_{\lvert z\rvert}^{l,3}$	$S_{\lvert z\rvert}^{l,4}$	$S_{\lvert z\rvert}^{l,5}$	$S_{\lvert z\rvert}^{l,6}$	$S_{\lvert z\rvert}^{l,7}$	$S_{\lvert z\rvert}^{l,8}$	$S_{\lvert z\rvert}^{l,9}$	$S_{\lvert z\rvert}^{l,10}$	$S_{\lvert z\rvert}^{l,11}$	$S_{\lvert z\rvert}^{l,12}$	$S_{\lvert z\rvert}^{l,13}$
$l = 1$	0.017	0.529	0.523	0.471	0.268	0.002	0.008	0.045	0.136	0.091	0.085	0.031	0.0001
$l = 2$	0.015	0.285	0.550	0.434	0.481	0.002	0.005	0.041	0.157	0.086	0.090	0.027	0.0001
$l = 3$	0.028	0.584	0.614	0.651	0.419	0.006	0.005	0.030	0.049	0.088	0.073	0.013	0.0001
$l = 4$	0.013	0.104	0.807	0.285	0.209	0.009	0.003	0.027	0.098	0.088	0.069	0.012	0.0001
$l = 5$	0.031	0.221	0.212	0.302	0.206	0.009	0.006	0.034	0.103	0.088	0.057	0.006	0.0001
	$A_{\lvert z\rvert}^{l,1}$	$A_{\lvert z\rvert}^{l,2}$	$A_{\lvert z\rvert}^{l,3}$	$A_{\lvert z\rvert}^{l,4}$	$A_{\lvert z\rvert}^{l,5}$	$A_{\lvert z\rvert}^{l,6}$	$A_{\lvert z\rvert}^{l,7}$	$A_{\lvert z\rvert}^{l,8}$	$A_{\lvert z\rvert}^{l,9}$	$A_{\lvert z\rvert}^{l,10}$	$A_{\lvert z\rvert}^{l,11}$	$A_{\lvert z\rvert}^{l,12}$	$A_{\lvert z\rvert}^{l,13}$
$l = 1$	0.342	0.671	0.759	0.405	0.163	0.042	0.013	0.248	0.243	0.243	0.412	0.199	0.042
$l = 2$	0.335	0.453	0.621	0.352	0.972	0.043	0.022	0.336	0.297	0.223	0.332	0.250	0.040
$l = 3$	0.494	0.579	0.316	0.567	0.309	0.041	0.016	0.216	0.225	0.235	0.359	0.197	0.049
$l = 4$	0.338	0.348	0.764	0.319	0.152	0.033	0.024	0.144	0.264	0.223	0.327	0.226	0.042
$l = 5$	0.138	0.595	0.637	0.238	0.142	0.032	0.024	0.218	0.273	0.222	0.308	0.184	0.042
	$S_C^{l,13}$	$A_C^{l,13}$											
$l = 1$	0.004	0.022											
$l = 2$	0.001	0.019											
$l = 3$	0.003	0.016											
$l = 4$	0.001	0.025											
$l = 5$	0.001	0.025											
	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_L^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$					
$C = 7$	0.006	0.009	0.012	0.01	0.027	0.018	0.027	0.029					
Rolling scheme	$S_{\lvert z\rvert}^{l,1}$	$S_{\lvert z\rvert}^{l,2}$	$S_{\lvert z\rvert}^{l,3}$	$S_{\lvert z\rvert}^{l,4}$	$S_{\lvert z\rvert}^{l,5}$	$S_{\lvert z\rvert}^{l,6}$	$S_{\lvert z\rvert}^{l,7}$	$S_{\lvert z\rvert}^{l,8}$	$S_{\lvert z\rvert}^{l,9}$	$S_{\lvert z\rvert}^{l,10}$	$S_{\lvert z\rvert}^{l,11}$	$S_{\lvert z\rvert}^{l,12}$	$S_{\lvert z\rvert}^{l,13}$
$l = 1$	0.528	0.460	0.843	0.472	0.142	0.002	0.001	0.007	0.008	0.093	0.281	0.026	0.0002
$l = 2$	0.258	0.350	0.580	0.251	0.229	0.007	0.004	0.009	0.002	0.081	0.240	0.024	0.0003
$l = 3$	0.559	0.310	0.741	0.437	0.160	0.002	0.001	0.002	0.002	0.066	0.288	0.013	0.0003
$l = 4$	0.38	0.310	0.479	0.463	0.170	0.001	0.001	0.003	0.008	0.088	0.275	0.021	0.0002
$l = 5$	0.285	0.249	0.297	0.225	0.120	0.007	0.001	0.003	0.003	0.063	0.327	0.019	0.0002
	$A_{\lvert z\rvert}^{l,1}$	$A_{\lvert z\rvert}^{l,2}$	$A_{\lvert z\rvert}^{l,3}$	$A_{\lvert z\rvert}^{l,4}$	$A_{\lvert z\rvert}^{l,5}$	$A_{\lvert z\rvert}^{l,6}$	$A_{\lvert z\rvert}^{l,7}$	$A_{\lvert z\rvert}^{l,8}$	$A_{\lvert z\rvert}^{l,9}$	$A_{\lvert z\rvert}^{l,10}$	$A_{\lvert z\rvert}^{l,11}$	$A_{\lvert z\rvert}^{l,12}$	$A_{\lvert z\rvert}^{l,13}$
$l = 1$	0.789	0.801	0.853	0.656	0.143	0.001	0.023	0.033	0.032	0.145	0.297	0.271	0.024
$l = 2$	0.267	0.550	0.812	0.244	0.568	0.002	0.018	0.024	0.045	0.127	0.283	0.337	0.037
$l = 3$	0.751	0.918	0.923	0.723	0.144	0.007	0.015	0.032	0.047	0.114	0.288	0.240	0.023
$l = 4$	0.358	0.476	0.575	0.448	0.108	0.001	0.015	0.031	0.031	0.142	0.313	0.271	0.034
$l = 5$	0.305	0.655	0.576	0.251	0.129	0.009	0.018	0.031	0.039	0.130	0.286	0.288	0.033
	$S_C^{l,13}$	$A_C^{l,13}$											
$l = 1$	0.001	0.009											
$l = 2$	0.0003	0.004											
$l = 3$	0.0003	0.010											
$l = 4$	0.0003	0.005											
$l = 5$	0.002	0.004											
	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_L^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$					
$C = 7$	0.008	0.008	0.006	0.009	0.038	0.025	0.019	0.031					

Notes: 13 autocontours, $C = [0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99]$.
$S_{|z|}^{l,7}, A_{|z|}^{l,7}$ for $l = 1, 2, \ldots, 5$; 7 refers to the 50% autocontour.
$S_L^{l,7}, A_L^{l,7}$ stacking lags up to $l = 2, \ldots .5$ and considering the 50% autocontour.
$S_C^{l,13}$ and $A_C^{l,13}$ stacking all 13 autocontours for one lag $l = 1, 2, 3, 4, 5$.
500 bootstrap samples, $T = 649$, $R = 360$, $P = 289$, $m = 0.69$.

tercept and the standard deviation of the error are state-dependent, we also observe a marginal rejection (p-values of around 4%) with the $S_{|z|}$ test for the 1% autocontour. The overall message of these testing results is that a shift in the level α^{S_t} and a shift in the standard deviation σ^{S_t} (Model 6) may be enough to understand the instability of the Phillips curve, but the density forecast of inflation changes will be more robust to unstable times if we also let the coefficient that links inflation and unemployment be state-dependent (Model 3). We report the testing results for Models 3 and 6 in Table 11.

For the preferred model, Model 3, we compute the natural rate of unemployment (NAIRU), which is the ratio of (minus) the intercept term to the coefficient on unemployment. The estimated NAIRU is 7.52% in recession times (state 1), and 5.53% in expansions (state 2). The persistence of state 1 is $p_{11} = 0.96$, while that of state 2 is $p_{22} = 0.99$, meaning that recessions are shorter than expansions. The estimated variance of state 1 is 0.20, while that of state 2 is 0.04, meaning that recessions are more volatile than expansions.

Table 11
Bootstrapped p-values: non-linear Phillips curve (fixed scheme).

Model 3	$S_{\|z\|}^{l,1}$	$S_{\|z\|}^{l,2}$	$S_{\|z\|}^{l,3}$	$S_{\|z\|}^{l,4}$	$S_{\|z\|}^{l,5}$	$S_{\|z\|}^{l,6}$	$S_{\|z\|}^{l,7}$	$S_{\|z\|}^{l,8}$	$S_{\|z\|}^{l,9}$	$S_{\|z\|}^{l,10}$	$S_{\|z\|}^{l,11}$	$S_{\|z\|}^{l,12}$	$S_{\|z\|}^{l,13}$
$l=1$	0.452	0.662	0.454	0.406	0.466	0.694	0.702	0.814	0.788	0.570	0.410	0.254	0.296
$l=2$	0.428	0.552	0.218	0.464	0.436	0.678	0.652	0.816	0.684	0.550	0.410	0.276	0.296
$l=3$	0.476	0.784	0.454	0.424	0.466	0.660	0.602	0.874	0.782	0.582	0.430	0.206	0.296
$l=4$	0.656	0.738	0.676	0.550	0.584	0.720	0.807	0.884	0.820	0.584	0.386	0.224	0.296
$l=5$	0.652	0.538	0.422	0.444	0.536	0.780	0.836	0.844	0.842	0.578	0.416	0.244	0.296
	$A_{\|z\|}^{l,1}$	$A_{\|z\|}^{l,2}$	$A_{\|z\|}^{l,3}$	$A_{\|z\|}^{l,4}$	$A_{\|z\|}^{l,5}$	$A_{\|z\|}^{l,6}$	$A_{\|z\|}^{l,7}$	$A_{\|z\|}^{l,8}$	$A_{\|z\|}^{l,9}$	$A_{\|z\|}^{l,10}$	$A_{\|z\|}^{l,11}$	$A_{\|z\|}^{l,12}$	$A_{\|z\|}^{l,13}$
$l=1$	0.442	0.644	0.430	0.390	0.568	0.682	0.856	0.822	0.818	0.596	0.45	0.278	0.294
$l=2$	0.402	0.446	0.318	0.362	0.326	0.488	0.674	0.834	0.720	0.570	0.446	0.308	0.296
$l=3$	0.606	0.748	0.446	0.388	0.454	0.558	0.804	0.868	0.790	0.606	0.434	0.230	0.294
$l=4$	0.638	0.768	0.666	0.348	0.580	0.722	0.860	0.874	0.824	0.584	0.400	0.226	0.296
$l=5$	0.632	0.514	0.404	0.432	0.522	0.722	0.830	0.852	0.860	0.582	0.412	0.246	0.390
	$S_C^{l,13}$	$A_C^{l,13}$											
$l=1$	0.614	0.596											
$l=2$	0.480	0.454											
$l=3$	0.494	0.482											
$l=4$	0.510	0.478											
$l=5$	0.488	0.472											
	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_L^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$					
$C=7$	0.518	0.566	0.548	0.598	0.534	0.578	0.564	0.618					
Model 6	$S_{\|z\|}^{l,1}$	$S_{\|z\|}^{l,2}$	$S_{\|z\|}^{l,3}$	$S_{\|z\|}^{l,4}$	$S_{\|z\|}^{l,5}$	$S_{\|z\|}^{l,6}$	$S_{\|z\|}^{l,7}$	$S_{\|z\|}^{l,8}$	$S_{\|z\|}^{l,9}$	$S_{\|z\|}^{l,10}$	$S_{\|z\|}^{l,11}$	$S_{\|z\|}^{l,12}$	$S_{\|z\|}^{l,13}$
$l=1$	0.044	0.238	0.335	0.141	0.401	0.433	0.474	0.559	0.478	0.577	0.299	0.254	0.042
$l=2$	0.042	0.345	0.346	0.168	0.392	0.514	0.521	0.442	0.418	0.552	0.260	0.215	0.053
$l=3$	0.042	0.242	0.369	0.110	0.375	0.512	0.477	0.498	0.457	0.595	0.301	0.210	0.044
$l=4$	0.043	0.316	0.378	0.154	0.411	0.566	0.457	0.423	0.485	0.506	0.297	0.239	0.052
$l=5$	0.049	0.321	0.403	0.196	0.434	0.492	0.479	0.470	0.482	0.598	0.303	0.198	0.052
	$A_{\|z\|}^{l,1}$	$A_{\|z\|}^{l,2}$	$A_{\|z\|}^{l,3}$	$A_{\|z\|}^{l,4}$	$A_{\|z\|}^{l,5}$	$A_{\|z\|}^{l,6}$	$A_{\|z\|}^{l,7}$	$A_{\|z\|}^{l,8}$	$A_{\|z\|}^{l,9}$	$A_{\|z\|}^{l,10}$	$A_{\|z\|}^{l,11}$	$A_{\|z\|}^{l,12}$	$A_{\|z\|}^{l,13}$
$l=1$	0.076	0.375	0.431	0.221	0.397	0.383	0.545	0.495	0.572	0.493	0.319	0.264	0.115
$l=2$	0.092	0.231	0.403	0.207	0.381	0.387	0.560	0.449	0.562	0.404	0.351	0.263	0.102
$l=3$	0.072	0.191	0.352	0.220	0.390	0.335	0.583	0.621	0.553	0.448	0.425	0.263	0.102
$l=4$	0.066	0.230	0.402	0.255	0.391	0.336	0.518	0.418	0.554	0.493	0.320	0.264	0.113
$l=5$	0.072	0.338	0.429	0.212	0.416	0.382	0.581	0.523	0.570	0.511	0.315	0.264	0.104
	$S_C^{l,13}$	$A_C^{l,13}$											
$l=1$	0.366	0.480											
$l=2$	0.299	0.433											
$l=3$	0.323	0.449											
$l=4$	0.326	0.436											
$l=5$	0.351	0.446											
	$S_L^{2,7}$	$S_L^{3,7}$	$S_L^{4,7}$	$S_L^{5,7}$	$A_L^{2,7}$	$A_L^{3,7}$	$A_L^{4,7}$	$A_L^{5,7}$					
$C=7$	0.488	0.475	0.483	0.464	0.464	0.419	0.461	0.404					

Notes: 13 autocontours, $C = [0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99]$.
$S_{|z|}^{l,7}, A_{|z|}^{l,7}$ for $l = 1, 2, \ldots, 5$; 7 refers to the 50% autocontour.
$S_L^{l,7}, A_L^{l,7}$ stacking lags up to $l = 2, \ldots .5$ and considering the 50% autocontour.
$S_C^{l,13}$ and $A_C^{l,13}$ stacking all 13 autocontours for one lag $l = 1, 2, 3, 4, 5$.
500 bootstrap samples, $T = 649$, $R = 360$, $P = 289$, $m = 0.69$.

5. Conclusion

We have provided a battery of tests for assessing the stability of density forecasts over time, which have important advantages for the empirical researcher. These tests are nuisance-parameter free and their asymptotic distributions can be tabulated. If the tests reject the null hypothesis of a stable density forecast, the shapes of the empirical generalized autocontours can be visualized in order to extract information regarding the direction of the rejection. Their finite sample properties are superior regardless of the estimation scheme (fixed, rolling, or recursive). In some cases, the Ave-tests tend to have a slightly better size than the Sup-tests. Both types are more powerful for detecting breaks in the intercept and slope coefficients than for detecting breaks in the variance. The tests can also be applied easily to multivariate random processes of any dimension. We have compared our tests with those of Rossi and Sekhposyan (2013) and found that our G-ACR tests are more powerful across the DGPs considered in their work. As one application of the tests, we have analyzed the stability of the Phillips curve. A linear model is rejected strongly in favor of a non-linear specification that allows shifts in the mean and variance of the inflation changes, as well as in the coefficient linking inflation and unemployment. The break of the linear model occurs during the Great Moderation years.

Acknowledgments

We are grateful to the participants at the International Symposium in Forecasting, Rotterdam (Holland), and to the seminar participants at the University of Southern California, University of California, San Diego, Universidad Carlos III de Madrid (UC3M), Universidad Autónoma de Madrid, Universidad Complutense de Madrid, and Bank of Spain for useful comments. Gloria González-Rivera acknowledges the financial support of the 2015 Chair of Excellence UC3M/Banco de Santander, as well as the support of the Spanish Ministerio de Economía y Competitividad (grant ECO2012-32854) and the UC-Riverside Academic Senate grants.

Appendix A

Proof of Proposition 1. The indicator function I_t^{k,α_i} is a Bernoulli random variable with the following moments: $E(I_t^{k,\alpha_i}) = \alpha_i$, $Var(I_t^{k,\alpha_i}) = \alpha_i(1-\alpha_i)$, and covariance

$$r_h^{\alpha_i} \equiv cov(I_t^{k,\alpha_i}, I_{t-h}^{k,\alpha_i}) = \begin{cases} 0 & \text{if } h \neq k \\ \alpha_i^{3/2}(1-\alpha_i^{1/2}) & \text{if } h = k. \end{cases}$$

Since the indicator process is stationary and ergodic, $\widehat{\alpha}_i(J)$ satisfies the condition of global covariance stationarity required for the FLCT in Theorem 7.17 of White (2000). Since $J = [Ps]$, $s \in [m, 1]$, and $r - k = [Pm]$, as $P \to \infty$, we write

$$\begin{aligned} W_P(s) &\equiv \frac{\sqrt{(r-k)}(\widehat{\alpha}_i(J) - \alpha_i)}{\sigma_{k,i}} \\ &= \frac{\sum_{t=R+1+J-r+k}^{R+J} (I_t^{k,\alpha_i} - \alpha_i)}{\sqrt{r-k}\sigma_{k,i}} \\ &= \frac{\sum_{t=R+1}^{R+J} (I_t^{k,\alpha_i} - \alpha_i)}{\sqrt{(r-k)}\sigma_{k,i}} - \frac{\sum_{t=R+1}^{R+J-(r-k)} (I_t^{k,\alpha_i} - \alpha_i)}{\sqrt{(r-k)}\sigma_{k,i}} \\ &= \frac{\sqrt{P}}{\sqrt{Pm}} \frac{\sum_{t=R+1}^{R+[Ps]} (I_t^{k,\alpha_i} - \alpha_i)}{\sqrt{P}\sigma_{k,i}} \\ &\quad - \frac{\sqrt{P}}{\sqrt{Pm}} \frac{\sum_{t=R+1}^{R+[P(s-m)]} (I_t^{k,\alpha_i} - \alpha_i)}{\sqrt{P}\sigma_{k,i}} \\ &\xrightarrow{d} \frac{1}{\sqrt{m}}(W(s) - W(s-m)), \end{aligned}$$

where $W(\cdot)$ is the standard Brownian motion and the limiting distribution, in the last line, is a direct consequence of the FCLT. Finally, by the continuous mapping theorem, we have:

$$\begin{aligned} S_{|z|} &= \sup_J \left| \frac{\sqrt{r-k}\widehat{\alpha}_i(J)}{\sigma_{k,\alpha_i}} \right| \xrightarrow{d} \sup \frac{|W(s) - W(s-m)|}{\sqrt{m}} \\ A_{|z|} &= \frac{1}{P-r+1} \sum_{\underline{J}}^{\bar{J}} \left| \frac{\sqrt{r-k}\widehat{\alpha}_i(J)}{\sigma_{k,\alpha_i}} \right| \\ &\xrightarrow{d} \int_{\underline{s}}^{\bar{s}} \frac{|W(s) - W(s-m)|}{\sqrt{m}}, \end{aligned}$$

where $\underline{J} = [P\underline{s}]$ and $\bar{J} = [P\bar{s}]$.

Proof of Proposition 2. Let Ω_k be the variance–covariance matrix of $\mathbf{C_k}(J)$, whose typical element $\omega_{i,j}$ is calculated as

$$\begin{aligned} cov(c_{k,i}, c_{k,j}) &= cov(I_t^{k,\alpha_i}, I_t^{k,\alpha_j}) + cov(I_t^{k,\alpha_i}, I_{t-k}^{k,\alpha_j}) \\ &\quad + cov(I_{t-k}^{k,\alpha_i}, I_t^{k,\alpha_j}) + o(1). \end{aligned}$$

If $i = j$, $\omega_{i,i} = var(\sqrt{T-k}(\widehat{\alpha}_i - \alpha_i)) = \alpha_i(1-\alpha_i) + 2\alpha_i^{3/2}(1-\alpha_i^{1/2})$. If $i < j$, $\alpha_i < \alpha_j$, and

$$\begin{aligned} cov(I_t^{k,\alpha_i}, I_t^{k,\alpha_j}) &= E(I_t^{k,\alpha_i} \times I_t^{k,\alpha_j}) - \alpha_i \times \alpha_j = \alpha_i(1-\alpha_j) \\ cov(I_t^{k,\alpha_i}, I_{t-k}^{k,\alpha_j}) &= E(I_t^{k,\alpha_i} \times I_{t-k}^{k,\alpha_j}) - \alpha_i \times \alpha_j \\ &= \alpha_i \times \alpha_j^{1/2} - \alpha_i \times \alpha_j \\ cov(I_{t-k}^{k,\alpha_i}, I_t^{k,\alpha_j}) &= E(I_{t-k}^{k,\alpha_i} \times I_t^{k,\alpha_j}) - \alpha_i \times \alpha_j \\ &= \alpha_i \times \alpha_j^{1/2} - \alpha_i \times \alpha_j. \end{aligned}$$

If $i > j$, the above expressions hold by just switching the subindexes i and j.

Since the vector $\mathbf{C_k}(J)$ is globally stationary, we can invoke a multivariate FLCT (see Theorems 7.29 and 7.30 of White, 2000). By following the same arguments as in Proposition 1, as $P \to \infty$,

$$\begin{aligned} \mathbf{W}_P(s) &\equiv \Omega_k^{-1/2}\mathbf{C_k}(J) \\ &\xrightarrow{d} \frac{1}{\sqrt{m}}(\mathbf{W}(s) - \mathbf{W}(s-m)), \end{aligned}$$

where $\mathbf{W}(s)$ is a C-variate Brownian process. By the continuous mapping theorem, we have

$$\begin{aligned} S_C &= \sup_J \mathbf{C}_k(J)'\Omega_k^{-1}\mathbf{C}_k(J) \\ &\xrightarrow{d} \sup_{s\in[m,1]} \frac{(\mathbf{W}(s) - \mathbf{W}(s-m))'(\mathbf{W}(s) - \mathbf{W}(s-m))}{m} \\ A_C &= \frac{1}{P-r+1} \sum_{\underline{J}}^{\bar{J}} |\mathbf{C}_k(J)'\Omega_k^{-1}\mathbf{C}_k(J)| \\ &\xrightarrow{d} \int_{\underline{s}}^{\bar{s}} \frac{(\mathbf{W}(r) - \mathbf{W}(r-m))'(\mathbf{W}(r) - \mathbf{W}(r-m))}{m}. \end{aligned}$$

Proof of Proposition 3. Let Λ_{α_i} be the variance–covariance matrix of $\mathbf{L}_{\alpha_i}(J)$, whose typical element $\lambda_{j,k}$ is calculated as

$$\lambda_{j,k} = \begin{cases} \alpha_i(1-\alpha_i) + 2\alpha_i^{3/2}(1-\alpha_i^{1/2}) & \text{if } j = k \\ 4\alpha_i^{3/2}(1-\alpha_i^{1/2}) & \text{if } j \neq k. \end{cases}$$

Therefore, the vector $\mathbf{L}_{\alpha_i}(J)$ is globally stationary, and we invoke a multivariate FLCT (Theorems 7.29 and 7.30 of White, 2000). By following the same arguments as in Proposition 1, as $P \to \infty$,

$$\mathbf{W}_P(s) \equiv \Lambda_{\alpha_i}^{-1/2}\mathbf{L}_{\alpha_i}(J) \xrightarrow{d} \frac{1}{\sqrt{m}}(\mathbf{W}(s) - \mathbf{W}(s-m)),$$

where $\mathbf{W}(s)$ is a L-variate Brownian process. By the continuous mapping theorem, we have

$$S_L = \sup_J \mathbf{L}_{\alpha_i}(J)'\Lambda_{\alpha_i}^{-1}\mathbf{L}_{\alpha_i}(J) \xrightarrow{d} \sup_{s\in[m,1]} \frac{(\mathbf{W}(s)-\mathbf{W}(s-m))'(\mathbf{W}(s)-\mathbf{W}(s-m))}{m}$$

$$A_L = \frac{1}{P-r+1}\sum_{\underline{J}}^{\bar{J}} \mathbf{L}_{\alpha_i}(J)'\Lambda_{\alpha_i}^{-1}\mathbf{L}_{\alpha_i}(J) \xrightarrow{d} \int_{\underline{s}}^{\bar{s}} \frac{(\mathbf{W}(s)-\mathbf{W}(s-m))'(\mathbf{W}(s)-\mathbf{W}(s-m))}{m}.$$

Appendix B

Description of the parametric bootstrap

Suppose that we are interesting in calculating the size of the test. Consider a DGP, e.g., $y_t = \alpha + \beta_1 y_{t-1} + \beta_2 x_{t-1} + \sigma\varepsilon_t$, with population parameters $\alpha = 1.5$, $\beta_1 = 0.5$, $\beta_2 = 0.6$, $\sigma = 1$ and $\varepsilon_t \to N(0, 1)$. The information on x_t is predetermined, meaning that it is known in advance. Assume a sample size, e.g., $T = 750$, that is split as $R = 450$ (estimation sample) and $P = 300$ (prediction sample), and where $r = 100$ ($m = 1/3$) is the rolling subsample over the prediction sample. We run 1000 Monte Carlo replications, and 500 bootstrap samples within each Monte Carlo replication. The Monte Carlo/bootstrap procedure consists of the following steps:

1. One Monte Carlo replication: draw from the conditional density of ε_t, e.g., from $N(0, 1)$. With the population parameters inserted into the DGP, generate one Monte Carlo sample of observations $\{y_t\}$.

2. With the first replication of observations $\{y_t\}$, estimate the model parameters, e.g., $\hat{\theta} = (\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \hat{\sigma})$. Obtain the sequence of one-step-ahead density forecasts (i.e., $f_t(y_t|\Omega_{t-1}, \hat{\theta})$) and the corresponding PITs (i.e., $u_t = \int_{-\infty}^{y_t} f_t(\nu_t|\Omega_{t-1}, \hat{\theta})d\nu_t$) for the first subsample of size 100 in the prediction sample. Using these PITs, construct the tests $|z_1|$, C_1, and L_1. By rolling over the prediction sample one observation at a time, we now have a second subsample of size 100, and we again construct the one-step-ahead density forecasts, the corresponding PITs, and the tests $|z_2|$, C_2, and L_2. If we keep rolling the observations for $r = 100$ over $P = 300$, we will have 201 statistics, that is $\{|z_j|\}_{j=1}^{201}$, $\{C_j\}_{j=1}^{201}$, and $\{L_j\}_{j=1}^{201}$. Over these collections of tests, we then calculate the supremum (S) and the average (A) in order to obtain the following six statistics: $S_{|z|}$, S_C, S_L and $A_{|z|}$, A_C and A_L.

We take care of parameter uncertainty by implementing a bootstrap procedure within each of the Monte Carlo replications. We draw 500 bootstrap samples as follows:

3. The parameters were fixed to the estimated values in step 2, e.g., $\hat{\theta} = (\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \hat{\sigma})$, and draws were taken from the conditional density of ε_t (e.g., from $N(0, 1)$) to generate bootstrap samples of $\{y_t^b\}$. For each bootstrap sample, we proceed to estimate the parameters $\hat{\theta}^b = (\hat{\alpha}^b, \hat{\beta}_1^{\,b}, \hat{\beta}_2^{\,b}, \hat{\sigma}^b)$ again, and, as in step 2, we also construct density forecasts, PITs, and the collection of tests $\{|z_j|^b\}_{j=1}^{201}$, $\{C_j^b\}_{j=1}^{201}$, and $\{L_j^b\}_{j=1}^{201}$ in order to finally obtain the statistics $S_{|z|}^b$, S_C^b, S_L^b, $A_{|z|}^b$, A_C^b and A_L^b. We repeat this 500 times to obtain the empirical distribution (quantiles) of each of the six statistics. For instance, we could compute the bootstrap critical value or 95% percentile for each test over the 500 samples. Suppose that we call the 95% percentile $S_{|z|}^{b(95\%)}$ for the tests $\{S_{|z|}^b\}_{b=1}^{500}$, and similarly for the other statistics.

4. Given the empirical distribution (over the 500 bootstrap samples) of each statistic, we compare the values of the tests computed in step 2, i.e. $S_{|z|}$, S_C, S_L, $A_{|z|}$, A_C, and A_L, with the bootstrapped critical values obtained in step 3. For instance, if $S_{|z|}^{b(95\%)} \le S_{|z|}$, we count it as one, otherwise we count it as zero.

Observe that steps 2, 3, and 4 are carried out for one Monte Carlo replication. Thus,

5. Repeat steps 1, 2, 3, and 4 for 1000 Monte Carlo replications and record the number of ones (step 4). The size of the test will be the percentage of ones over these 1000 replications.

Appendix C. Supplementary material

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.ijforecast.2016.10.003.

References

Amisano, G., & Giacomini, R. (2007). Comparing density forecast via weighted likelihood ratio tests. *Journal of Business and Economic Statistics*, *25*(2), 177–190.

Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, *61*(4), 821–856.

Andrews, D. W., & Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica*, *62*(6), 1383–1414.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, *28*, 591–605.

Corradi, V., & Swanson, N. R. (2006). Predictive density evaluation. In G. Elliott, C. W. J. Granger, & A. Timmermann (Eds.), *Handbook of economic forecasting*: *vol. 1* (pp. 187–228). North-Holland.

Diebold, F., Gunther, A., & Tay, A. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, *39*(4), 863–883.

González-Rivera, G., Senyuz, Z., & Yoldas, E. (2011). Autocontours: dynamic specification testing. *Journal of Business and Economic Statistics*, *29*(1), 186–200.

González-Rivera, G., & Sun, Y. (2015). Generalized autocontours: evaluation of multivariate density models. *International Journal of Forecasting*, *31*(3), 799–814.

González-Rivera, G., & Yoldas, E. (2012). Autocontour-based evaluation of multivariate predictive densities. *International Journal of Forecasting*, *28*(2), 328–342.
Goyal, A., & Welch, I. (2003). Predicting the equity premium with dividend ratios. *Management Science*, *49*, 639–654.
Paye, B., & Timmermann, A. (2006). Instability of return prediction models. *Journal of Empirical Finance*, *13*(3), 274–315.
Pesaran, M. H., & Timmermann, A. (2002). Market timing and return prediction under model instability. *Journal of Empirical Finance*, *9*(5), 495–510.
Rapach, D. E., & Zhou, D. (2014). Forecasting stock returns. *Handbook of Economic Forecasting*, *2*, 328–383.
Rogoff, K.S., & Stavrakeva, V. 2008. The continuing puzzle of short horizon exchange rate forecasting. NBER Working Papers 14071.
Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, *23*(3), 470–472.
Rossi, B. (2006). Are exchange rates really random walks? Some evidence robust to parameter instability. *Macroeconomic Dynamics*, *10*(1), 20–38.
Rossi, B. (2014). Advances in forecasting under instability. *Handbook of Economic Forecasting*, *2*, 1203–1324.
Rossi, B., & Sekhposyan, T. (2010). Have economic models' forecasting performance for US output growth and inflation changed over time, and when? *International Journal of Forecasting*, *26*(4), 808–835.
Rossi, B., & Sekhposyan, T. (2013). Conditional predictive density evaluation in the presence of instabilities. *Journal of Econometrics*, *177*, 199–212.
Stock, J. H., & Watson, M. W. (1999). Forecasting inflation. *Journal of Monetary Economics*, *44*, 293–335.
Stock, J. H., & Watson, M. W. (2003). Forecasting output and inflation: the role of asset prices. *Journal of Economic Literature*, *41*, 788–829.
White, H. (2000). *Asymptotic theory for econometricians* (revised ed.). Academic Press.

Gloria González-Rivera is Professor of Economics at the University of California, Riverside.

Yingying Sun was Assistant Professor of Economics at Huazhong University of Science and Technology, China, she is currently at Bank of the West.